UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA VEGETAL

Assessment of genetic variability in the exotic invasive species *Acacia longifolia* using molecular markers

Sara Filipa da Silva Lucas Sousa Vicente

Mestrado em Biologia Molecular e Genética

Dissertação orientada por:
Professora Doutora Maria Helena Machado Trindade de Donato e Professora Doutora Cristina Maria Filipe Máguas da Silva Hanson

2016

This thesis resulted in the publication of a poster:

Sara Vicente, Helena Trindade, Octávio Paulo, Miguel Prado, Cristina Máguas (2016). ***Genetic Variability Assessment of the Exotic Invasive Species* Acacia longifolia *Along Portuguese Coastal Sand Dunes Using Molecular Markers: a Preliminary Study.*** International Symposium on Biological Invasion II Jornada Botânica, 25 and 26 of July 2016, Minas Gerais, Brazil.

Acknowledgements

É com o maior orgulho que digo que durante este ano aprendi imenso e fui muito, muito feliz. Contudo, apesar de serem minhas as palavras desta dissertação, a realização deste projeto passou não só por mim mas pela ajuda e apoio de várias outras pessoas maravilhosas, as quais tenho muito gosto em destacar nos próximos parágrafos.

Em primeiro lugar, gostaria de agradecer às minhas orientadoras por quem tenho muito carinho, as Professoras Doutoras Helena Trindade e Cristina Máguas, pela confiança que depositaram em mim e por todo o apoio durante a realização deste projeto. À professora Helena Trindade, obrigada pela paciência e pela disponibilidade de me acompanhar e ensinar a trabalhar num laboratório. À professora Cristina Máguas, obrigada por todo o conhecimento sobre as acácias e o trabalho de campo que me passou. Obrigada por tudo o que ambas me ensinaram dentro e fora do laboratório. Obrigada por me ensinarem a ser cientista.

Muito obrigada ao Professor Doutor Octávio Paulo, cujo conhecimento foi imprescindível para a análise dos microssatélites e para os cálculos dos parâmetros populacionais. Sem a sua ajuda grande parte dos resultados do projeto não seriam possíveis, e eu estou muito grata pela sua disponibilidade.

Um especial obrigada ao Miguel Prado por ter disponibilizado as acácias da sua herdade em Vila Nova de Milfontes ao meu projeto, pela ajuda na recolha das amostras e pela informação histórica sobre esta espécie em Portugal que me forneceu. Obrigada pelo entusiasmo e interesse que mostrou perante o meu projeto.

À Márcia Vieira, Andreia Anjos, Patrícia Fernandes e Catarina Costa, obrigada pelo apoio nas recolhas das amostras. Nunca teria conseguido recolher todas as minhas acácias sem a vossa ajuda e disponibilidade. Obrigada Teresa Mexia pela ajuda com os dados do GPS.

Aos colegas do laboratório 2.1.17 e do gabinete 2.1.49 da FCUL, obrigada por me terem recebido tão bem e por toda a ajuda e apoio que me deram durante este ano.

A todos os meus amigos, que sempre tiveram palavras de apoio e coragem para me dar e desejaram o meu sucesso, obrigada. Obrigada em especial à Rafaela Heitor, melhor amiga há 10 anos, por sempre ter ouvido as minhas frustrações e problemas enquanto desenvolvia este projeto, apesar de trabalhar noutra área e nem sempre perceber do que eu estava a falar (não fizeste mais que a tua obrigação!). És das melhores pessoas que conheço, e espero vir a poder fazer o mesmo por ti! Obrigada à Rita Beja e à Maria Inês Catarino, amigas de longa data e do coração, por sempre arranjarem tempo para mim nas vossas vidas ocupadas. Obrigada à Ana Russo pelos anos de amizade. Obrigada à minha bioquímica favorita e companheira desde o início dos anos de faculdade, Joana Andrade – apesar de estares longe, não me esqueço de ti! Obrigada por todo o teu apoio, a qualquer hora e em qualquer lugar.

A toda a minha família, obrigada pelas palavras de apoio e coragem. Em especial, muito obrigada aos meus pais, Manuela e Ramiro Vicente, e à minha tia (e fada madrinha) Isabel Lucas, por tudo aquilo que fizeram e ainda hoje fazem por mim, e pelo apoio e incentivo incondicionais em todas as decisões que tomei até hoje. Obrigada por acreditarem em mim. Sem vocês nada disto seria possível, pelo que vos dedico esta dissertação.

Finalmente, quero dizer simplesmente obrigada por ter a enorme sorte de fazer o que mais gosto e por ter tantas pessoas a quem agradecer.

Resumo

Acacia longifolia (Andrews) Willd. é uma leguminosa arbórea com capacidade para fixar azoto atmosférico, proveniente do sudoeste da Austrália e da Tasmânia. Conhecida como "acácia-de-espigas" devido à sua característica flor amarela alongada, esta espécie foi trazida para Portugal pelos serviços florestais de forma a conservar e fixar as dunas da Costa Portuguesa. A sua plantação começou no final do século XIX e estendeu-se a vários locais do país, como as dunas de São Jacinto (a norte de Aveiro), o sistema dunar de Quiaios-Mira (Figueira da Foz) e a costa Vicentina, nomeadamente Vila Nova de Milfontes (Odemira). Contudo, *A. longifolia* rapidamente se espalhou pelo território Português, competindo com as plantas nativas por recursos, tais como água, luz solar e nutrientes. O sucesso desta espécie foi tal que atualmente constitui uma ameaça à diversidade das espécies nativas e ao funcionamento dos ecossistemas, o que levou à classificação de *A. longifolia* como invasora em Portugal. Vários estudos têm sido realizados nas áreas da ecologia, fisiologia e fenologia, os quais demonstraram que indivíduos desta espécie têm um comportamento diferente consoante o ambiente em que estão inseridos, evidenciando uma grande plasticidade e adaptação, características que poderão ser potenciadoras do processo invasor. No entanto, pouco se sabe sobre esta espécie ao nível molecular, e em particular sobre a sua diversidade genética.

Neste trabalho, tínhamos como objectivo estudar a diversidade genética de *A. longifolia* em diferentes regiões da costa Portuguesa com um clima e uma gestão florestal diferentes, bem com na região de Vila nova de Milfontes onde pelo registo histórico foi possível identificar os indivíduos originalmente plantados entre 1960/1970 pelos serviços florestais. Para tal, foram recolhidos filódios de *A. longifolia* em três locais: 25 amostras em Osso da Baleia (Pombal), 25 em Pinheiro da Cruz (Setúbal) e 39 em Vila Nova de Milfontes, onde neste último se procedeu a um estudo mais pormenorizado que incluiu 14 novas amostras. O estudo genético teve como base dois tipos de marcadores moleculares: ISSRs (*inter-simple sequence repeats*) e SSRs (*simple sequence repeats*) ou microssatélites. O DNA foi extraído a partir dos filódios pelo método de CTAB (brometo de cetiltrimetilamonio) e precipitado com etanol frio. Após confirmação da qualidade do DNA, foram avaliados os polimorfismos moleculares através de PCR (*Polymerase Chain Reaction*) pela amplificação iniciada com diversos *primers* (iniciadores) de ISSRs e microssatélites. Foram testados 31 *primers* de ISSRs, sendo 16 selecionados para o estudo nos três locais do país e 11 selecionados para o estudo detalhado de Vila Nova de Milfontes. Com os microssatélites, testaram-se 11 pares de *primers*, sendo selecionados 3 para a análise final: os que amplificavam os loci DCLOC, CPUH4 e APZIZ. No caso dos ISSRs, os resultados foram visualizados em gel de agarose 2%, enquanto que para os microssatélites os *primers* foram marcados com sondas fluorescentes. Após amplificação por PCR,

a electroforese capilar (STAB Vida) gerou resultados finais na forma de electroferogramas. A análise dos resultados originou matrizes binárias (preenchidas com 0 e 1) para os ISSRs e matrizes de alelos amplificados para os microssatélites, que foram posteriormente analisadas com *softwares* de forma a calcular os parâmetros descritivos e o índice de fixação de Wright (F_{ST}) para comparação entre os locais de recolha das amostras.

A análise com ISSRs das três localidades do país gerou 275 bandas no total, das quais 258 são polimórifcas (94%), enquanto que a análise de Vila Nova de Milfontes gerou 137 bandas, todas elas polimórficas. O dendrograma obtido a partir dos resultados de ISSRs considerando as três regiões geográficas demonstra que as amostras têm tendência a agrupar-se de acordo com o seu local de recolha, com exceção de um grupo de 9 indivíduos de Osso da Baleia que se agrupa juntamente com Vila Nova de Milfontes. No entanto, as separações dos grupos ocorrem todas com coeficientes de similaridade de Dice muito próximos e relativamente elevados (S_{Dice} entre 0.63 e 0.85), indicando um elevado nível de similaridade da *A. longifolia* entre os três locais em estudo apesar das distâncias que os separam. O mesmo ocorre quando se avalia o dendrograma obtido apenas com amostras Vila Nova de Milfontes (S_{Dice} entre 0.61 e 0.79), com a exceção que estas últimas não se agrupam de acordo com o local de recolha. Os resultados com ISSRs demonstram ainda que há uma diferenciação genética baixa mas estatisticamente significativa (F_{ST} = 0.070) entre amostras de Vila Nova de Milfontes, Osso da Baleia e Pinheiro da Cruz, o que possivelmente aponta para uma origem comum de *A. longifolia* em Portugal. No entanto, as amostras de Pinheiro da Cruz mostraram ser as mais diferentes dos restantes locais, enquanto que as amostras de Osso da Baleia revelaram-se muito semelhantes às amostras de Vila Nova de Milfontes, apesar da distância de cerca de 260 Km que as separa. Já em Vila Nova de Milfontes, amostras recolhidas em várias zonas desta localidade mostraram ter uma diferenciação genética estatisticamente significativa muito inferior à obtida quando se consideram os três locais do país (F_{ST} = 0.013), o que aponta para que as acácias plantadas pelos serviços florestais se tenham espalhado por esta localidade. No entanto, as diferenças observadas entre as plantas originalmente plantadas pelos serviços florestais e as restantes nesta localidade indicam que já há alguma variabilidade que terá ocorrido num período de cerca de 50 anos. A análise de microssatélites gerou no total 7 alelos a partir de dois pares de *primers* (3 alelos do locus DCLOC e 4 alelos do locus CPUH4), tendo-se excluído os resultados do locus APZIZ por apresentarem mais do que dois alelos por indivíduo. Os microssatélites não demonstram uma diferenciação genética estatisticamente significativa, mas este resultado é possivelmente consequência de se terem analisado apenas dois loci. Recorrendo-se a uma análise mais simplista, os alelos foram classificados como comuns (C) ou raros (R), sendo um alelo raro aquele que apresenta no geral baixa frequência e está ausente em pelo menos uma localidade. A análise das frequências alélicas demonstra que todas as localidades partilham alelos raros mas com frequências distintas, o

que apoia a hipótese de uma origem comum de *A. longifolia* em Portugal. Amostras recolhidas na localidade de Vila Nova de Milfontes demonstram ter frequências de alelos raros semelhantes entre si (entre 20–40%), apoiando a hipótese de que as acácias plantadas pelos serviços florestais se terão espalhado por esta localidade, enquanto que Osso da Baleia é a localidade com maior frequência de alelos raros (54%). Pinheiro da Cruz apresenta apenas 9% de alelos raros, o que confirma que esta localização é a mais distinta de todas as estudadas.

A análise conjunta dos resultados obtidos pelos dois marcadores moleculares levanta a hipótese de uma origem comum de *A. longifolia* em Portugal, resultante de uma introdução única – a plantação pelos serviços florestais – e os indivíduos desta espécie serão, portanto, semelhantes ao nível molecular. É plausível considerar-se que os serviços florestais terão utilizado o mesmo lote de sementes para as plantações ao longo do país, podendo ter ocorrido transferência de sementes de uma localização para as outras. Assim, as acácias situadas ao longo de Vila Nova de Milfontes tiveram a sua origem nas que foram plantadas pelos serviços florestais junto à costa. Não tivemos acesso a nenhum registo de plantação de acácias em Osso da Baleia, mas é plausível considerar que as acácias do sistema dunar de Quiaios-Mira se tenham espalhado em direção a Sul, tendo chegado eventualmente a Osso da Baleia, que se localiza apenas a 25 Km de distância. Relativamente a Pinheiro da Cruz também não houve acesso a qualquer acesso a informação histórica, mas é possível que tanto as acácias plantadas em Vila Nova de Milfontes como as plantadas na Costa da Caparica se possam ter disseminado até lá, uma vez que ambas as localizações distam cerca de 60 Km de Pinheiro da Cruz, e que a frequência dos alelos raros tenha diminuído ao longo do processo invasivo. É de notar que a população de acácias deste local se encontra mais isolada, pelo facto de se tratar de um estabelecimento prisional, contribuindo para o declínio da frequência dos alelos raros. Contudo, os resultados obtidos com microssatélites são preliminares, pelo que de futuro serão necessários estudos com um maior número de loci e igualmente com outros tipos de marcadores moleculares, que darão novas informações e ajudarão a esclarecer a distribuição desta espécie em Portugal.

Acacia longifolia é uma espécie que exibe um comportamento extremamente invasor contribuindo para a perda de diversidade das espécies nativas. Como tal, esclarecer a origem, o padrão de distribuição e a diversidade genética desta espécie em Portugal é importante para a compreensão do seu mecanismo e capacidade invasores de forma a promover a conservação das espécies nativas que possuem valor económico e biológico.

Palavras-chave: *Acacia longifolia*, invasora, ISSRs, SSRs, diversidade genética.

Abstract

Acacia longifolia (Andrews) Willd. is a nitrogen fixing tree or shrub original from the Southeast Australia and Tasmania. Also known as "Sydney Golden Wattle", it was brought to Portugal by the forestry services to promote the conservation of sand dunes and introduced in various locations along the Portuguese coastal line, including the São Jacinto sand dunes (north of Aveiro), Costa da Caparica (Almada, Setúbal), the Quiaios-Mira sand dunes (Figueira da Foz) and Vila Nova de Milfontes (Odemira). However, *A. longifolia* quickly spread and dominated nearby soils, eventually replacing native species. This species is so successful that it is considered a serious environmental problem, leading to its classification as invasive in Portugal.

A. longifolia has been subjected to many studies in different fields such as ecology and land management, but not much is known about this species on a molecular level. In this study, we assessed the genetic diversity of 89 accessions from three locations in the Portuguese coastal line – 25 from Osso da Baleia (Pombal), 25 from Pinheiro da Cruz (Setúbal) and 39 from Vila Nova de Milfontes – using two types of molecular markers: inter-simple sequence repeats (ISSRs) (16 primers) and simple-sequence repeats (SSRs) or microsatellites (2 primers). We also conducted a more detailed study on the genetic diversity of this species in Vila Nova de Milfontes based on 53 accessions (11 ISSR primers only and 2 microsatellite primers), collecting samples from different sites including the region where acacias were introduced by the forestry services. ISSR results revealed that samples from the three different locations in the coastal line show low but significant molecular differentiation (F_{ST} = 0.070), resulting in their clustering by collection region, while samples from different sites in Vila Nova de Milfontes have much lower genetic differentiation (F_{ST} = 0.013) and show no particular clustering. Microsatellite analysis evidenced that samples from all location share the same alleles, indicating the possibility of a single origin of *A. longifolia* in Portugal. We hypothesize that the genetic similarity of acacias found in different regions of the Portuguese coast derives from the fact that the forestry services used the same allotment of seeds. Another possibility relies on the transference of seeds from one location to another, which in turn spread and invaded adjacent soils. Further studies with molecular markers are needed in the future to better understand the distribution of this species in Portugal.

Keywords: *Acacia longifolia*, invasive, ISSRs, SSRs, genetic diversity.

Table of Contents

List of Tables

List of Figures

Abbreviations

A	number of detected alleles
A_e	number of effective alleles
bp	molecular weight units (base pairs)
BSA	Bovine Serum Albumin
CTAB	cetyl trimethylammonium bromide
dNTPs	Deoxynucleotides; includes dATP, dCTP, dGTP, dTTP
EDTA	ethylenediaminetetraacetic acid
F_{ST}	Wright's fixation index
F_{IS}	Inbreeding coefficient
h	Nei's gene diversity
H_e	expected heterozygosity
H_o	observed heterozygosity
H_S	mean expected heterozygosity within subpopulations
H_T	mean expected heterozygosity in total population
HWE	Hardy-Weinberg equilibrium
I	Shannon's information index
ISSRs	Inter-Simple Sequence Repeats
MFAnt	Samples from the plantation site in Vila Nova de Milfontes
MFCL	Samples from the Vila Nova de Milfontes' coastal line
MFRec	Samples from a recently invaded site in Vila Nova de Milfontes
ns	non-significant
OB	Samples from Osso da Baleia
PC	Samples from Pinheiro da Cruz
PVP	polyvinylpyrrolidone
RAPDs	Random Amplified Polymorphic DNA
RFLPs	Restriction Fragment Length Polymorphisms
SE	standard error
SSRs	Simple Sequence Repeats
TAE	Tris-Acetate-EDTA
TE	Tris-EDTA
Tris-HCl	Tris-hydrochloride
U	enzymatic units
UPGMA	Unweighted Pair Group Method with Arithmetical Averages
UV	Ultra-violet
VNTRs	Variable Number of Tandem Repeats

1. Introduction

Many exotic species are introduced deliberately in a given environment because of economic or industrial interest[1]. Once in a new environment, the introduced species might have some advantages when compared to the indigenous species, like the absence of natural enemies[1], better environment resistance[1] or the ability to establish symbiotic relations with the soil microorganisms[2]. These advantages enable the introduced species to grow and spread faster, effectively invading the soil and outcompeting the indigenous species for nutrients, sunlight and water, ultimately replacing them[1]. Many plants exhibit this invasive behavior, including members of the genus *Acacia*, which are considered one of the most aggressive invaders worldwide[3]. In Portugal, *Acacia longifolia* in particular has become a serious environmental problem along the coastal area, changing sand dune habitats and coastal forests.

1.1. *Acacia longifolia*

Acacia longifolia (Andrews) Willd., also known as "Sydney Golden Wattle", is a nitrogen fixing plant from the southeast Australia and Tasmania. It is a small tree or shrub with elongated leaves and yellow flowers (see Figure 1.1). This species is a member of the Mimosoideae subfamily included in the Fabaceae family, also known as Leguminosae or "pea family".

Due to its amazing invasive capacity, this species has been subjected to many studies to understand its invasive mechanisms in order to prevent its further invasion of soils. It is known that the invasive capacity of *A. longifolia* is due to its ability to form a high number of root nodules in different soils[2] and the establishment of a symbiotic relationship with the soil's mutualisms[2], enabling spreading to nutrient-poor soils (like the dune systems). Also, due to its high number of seeds and growth rate, *A. longifolia* has the ability to control the available sunlight to other plants, eventually outcompeting them and occupying their space[4]. Furthermore, this species is capable of quick regeneration after fires, because its seeds are resistant and germination is promoted by the fire itself[4]. This is an important mechanism that probably contributed to its invasiveness in Portugal, a country with frequent fires during the summer.

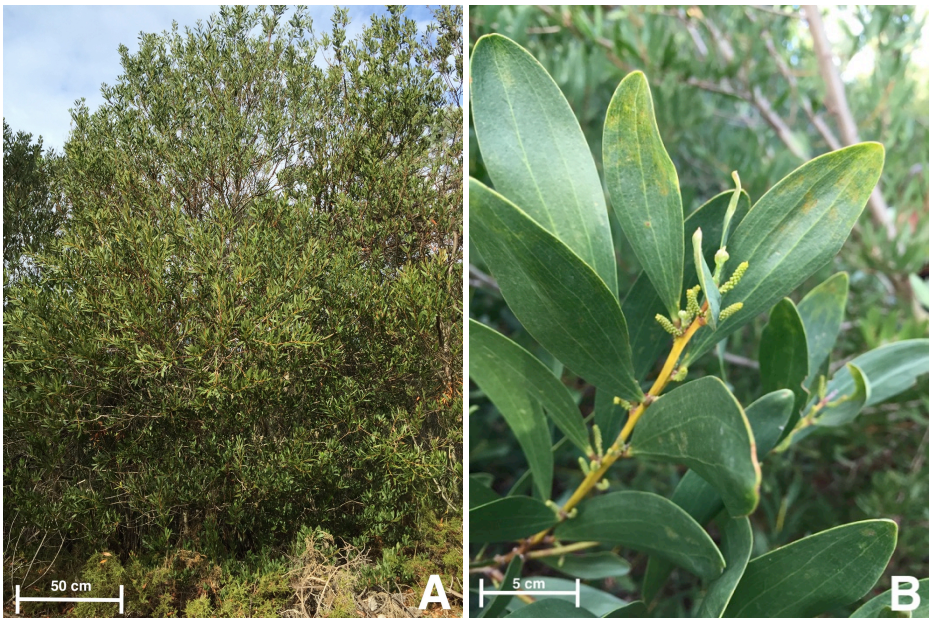

Figure 1.1: *Acacia longifolia* growing in the field. A – *A. longifolia* tree. B – Close up of an *A. longifolia* branch, showing the elongated phyllodes and the green rod shaped parts that are the flower buds.

Acacia longifolia was introduced in the Portuguese coastlines by the forestry services at different periods in the late 19th century/early 20th century as a way to prevent the erosion of sand dunes[5]. Records show that between 1888 and 1929, *A. longifolia* was planted in São Jacinto (north of Aveiro) to fixate sand dunes that protected plantations of *Pinus pinaster* and *Myrica faya*[6], and again in 1906 in Costa da Caparica (Almada, Setúbal)[7]. In 1924, to repopulate the sand dunes in Quiaios-Mira (Figueira da Foz) left bare after the French invasions, *A. longifolia* was also planted by the forestry services[6]. Later on, in 1948, *A. longifolia* was introduced once again in the Quiaios-Mira dunes due to the environmental conditions[6]. This species was also introduced in the sand dunes of Vila Nova de Milfontes in the late 1960s/early 1970s (Miguel Prado, personal communication).

Since its introduction in Portugal, *A. longifolia* has invaded many adjacent environments including *P. pinaster* plantations, which have a high economical impact[8]. The invasive success of *A. longifolia* in Portugal is such that, even introduced only roughly 130 years ago, it can be found throughout the country and is negatively impacting the native species' diversity[4] and the functioning of ecosystems[9]. A recent study on the phenology and reproductive success of *A. longifolia* has taken place in two different locations of Portugal – Osso da Baleia and Pinheiro da Cruz – and the results showed strong differences between the two sites, namely in the reproductive success, which is higher in Osso da Baleia, and in the phenological timings, which occurred earlier in Pinheiro da Cruz than in Osso da Baleia, in particular the fruit growth and ripening and the reproductive phenophases[8]. Also, acacias from Pinheiro da Cruz show a shorter flowering period when compared to those from Osso da Baleia[8]. These results show a clear influence of

environmental factors in the phenology of *A. longifolia*, since Osso da Baleia is classified as a mesic habitat (higher moisture) while Pinheiro da Cruz is classified as a xeric habitat (lower moisture)[8].

Despite all these studies, not much is known about *A. longifolia* on a molecular level. In recent years, the interest in molecular studies for invasive species has increased due to their importance in this field. One subject that is the target to many studies is the influence of the genome size in the invasiveness capacity of a species. In 2010, Lavergne *et al.*[10] found that invasive *Phalaris arundinacea* had smaller genome sizes when compared to non-invasive, and that this phenomenon is probably due to the natural selection of smaller genomes during the invasive process. Later, in 2014, Pandit *et al.*[11] performed a larger study that included 890 species from 63 genera and found that genome size correlates negatively with invasiveness, while ploidy level correlated positively. Interestingly, Gallagher *et al.*[12] performed this type of study in Australian species of *Acacia* – including *A. longifolia* – and found no correlation between genome size and invasiveness. However, invasive acacias tended to be taller and had a larger native range distribution. All these results stated above clearly show that incorporating genome size and other molecular analysis should elucidate invasive mechanisms. In this study, we performed an assessment of the genetic variability of *A. longifolia* in three different regions of the Portuguese coast line using two types of molecular markers: Inter-Simple Sequence Repeats (ISSRs) and Simple Sequence Repeats (SSRs) or microsatellites. Such analysis is important to better understand the invasion patterns of this species, and thus to better control and prevent further invasions in Portugal and other locations.

1.2. Molecular Markers

Genetic markers are features located in specific regions of a chromosome that allow distinction between individuals, populations or species, and are considered representative of changes in the genome. Genetic markers consist of two types: morphological and molecular[13]. Morphological markers have been widely used since they are easily monitored by visual analysis. However, they have several limitations, including their limited number and alterations by environmental, epistatic and pleiotropic interactions[14], which turned researchers' attention towards molecular markers in recent years. Molecular markers are not affected by environmental interactions, epistatic or pleiotropic effects, and can be used regardless of the type of tissue[13]. They can be classified as biochemical (isozymes) or DNA markers[13], and since isozymes also have some limitations, many efforts are made to develop new and more precise DNA markers. DNA markers can be divided into two categories: hybridization-based markers, like Restriction Fragment Length Polymorphism (RFLP) and Variable Number of Tandem Repeats (VNTRs); and PCR-based markers, like Random Amplified Polymorphic DNA (RAPDs), Inter-Simple Sequence Repeats (ISSRs) and Simple Sequence Repeats (SSRs)[14].

1.2.1. Inter-Simple Sequence Repeats (ISSRs)

Inter-simple sequence repeats (ISSRs) are highly polymorphic dominant molecular markers used mainly to study plant genetic diversity, described for the first time in 1994[15]. ISSRs are amplified by a polymerase chain reaction (PCR) in the presence of primers consisting of repeated sequences of DNA motifs (for example $(AG)_n$ or $(TC)_n$)[16]. In other words, ISSR primers contain microsatellite sequences that will anneal to complementary regions in the genome. The PCR reaction amplifies the region between two oppositely oriented adjacent microsatellites [15,17], originating amplification products that can be polymorphic. Therefore, no previous knowledge of the sequence is needed[18] and the same primer sequence function as both the forward and reverse primer, making this technique simple, low cost and with high level of polymorphism[19] and throughput. A diagram of the amplification process with an ISSR primer is shown on Figure 1.2.

ISSR markers have been widely used in many studies of genetic variability of populations and cultivars of many species in Portugal. An example is the assessment of the genetic diversity in Azorean *Pittosporum undulatum*[20], an invasive species from Australia, using ISSR markers and volatile analysis. *Acacia* species have already been studied with molecular markers: an example is the assessment of genetic diversity in *A. senegal* from Kenya using both RAPD and ISSR markers[21]. However, for A. *longifolia* no previous reports using molecular markers are available on the literature.

1.2.2. Simple Sequence Repeats (SSRs) or Microsatellites

Simple sequence repeats (SSRs) or microsatellites are composed of 5 to 40 repeats of tandem DNA motifs ranging between 1 to 6 nucleotides in length[22]. SSRs are found in high frequency in most taxa and can be used as co-dominant markers[23] to infer variability among individuals of the same species by performing a PCR reaction[22]. Microsatellites are highly informative and, since they are co-dominant, allow distinction between homozygotes and heterozygotes. However, non-repetitive DNA flanks SSRs and thus previous knowledge of the DNA sequence is needed in order to design primers for the PCR reaction[18]. Also, when using an automated method like capillary electrophoresis for fragment analysis, the primers must be labeled with fluorescent dyes, which make this technique rather costly. Figure 1.2 is an example of the PCR process with microsatellite primers.

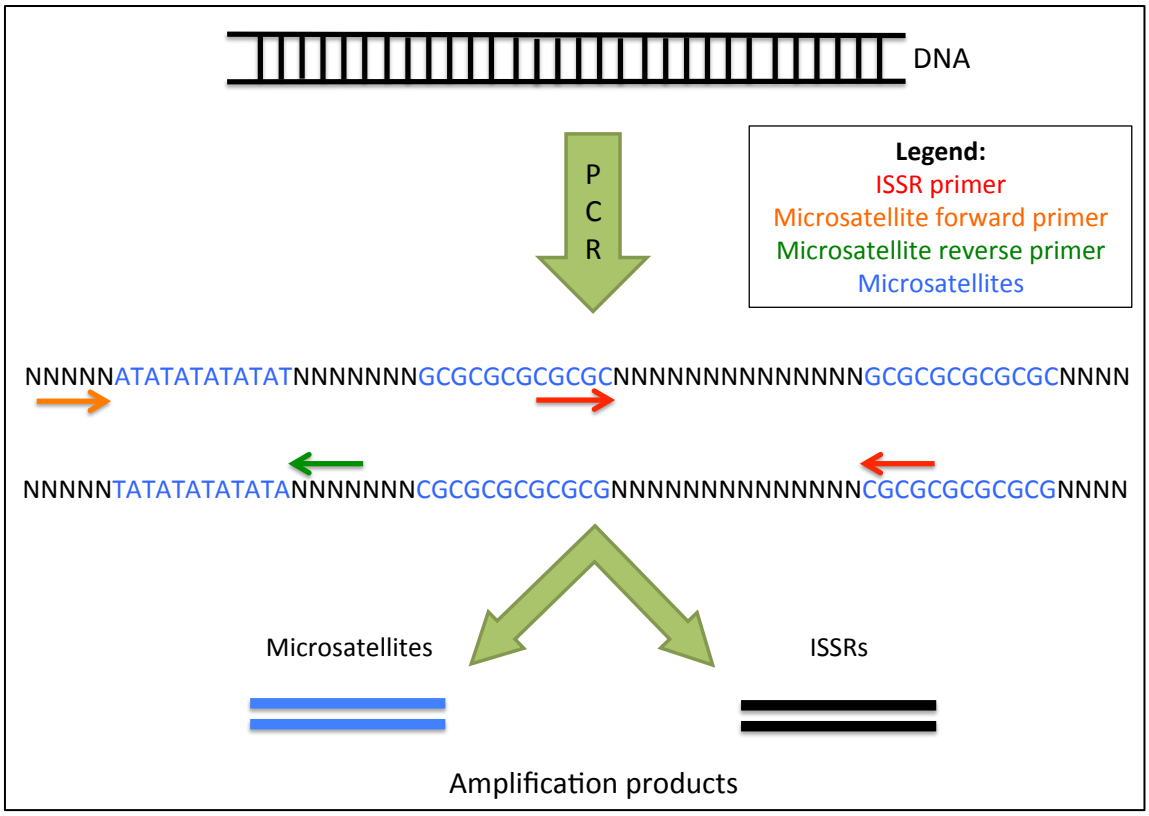

Figure 1.2: Example of the PCR amplification at a locus with ISSR and microsatellite primers. During the annealing process, the ISSR primer (red) attaches to two adjacent microsatellites (sequence in blue as an example), resulting in the amplification of the region between them. Different fragments of different sizes will be amplified as a result of primer annealing throughout the genome. The microsatellite technique relies on the attachment of a forward primer (orange) and reverse (green) to a known region of the genome; the primers attach themselves to regions adjacent to the microsatellites, allowing the primed amplification of microsatellites regions in between.

The SSRs analysis is very popular nowadays for assessments of genetic diversity, and population studies using this marker have already been done in many species of *Acacia* spp. like *A. dealbata*[24], *A. koa*[25] and *A. senegal*[26]. Also, cross-species studies have taken place where SSRs primers designed for one species of Acacia are used on other species within the same genus, and some have yielded positive results including studies with vulnerable Acacias from Australia's arid zone[27,28]. However, there are no studies with *A. longifolia*, and thus no SSR primers for this species have been described yet.

1.3. Population Studies and Descriptive Parameters: Wright's F-statistics

Population subdivision is an occurring phenomenon in natural populations, and it usually leads to reduction of heterozygosity due to inbreeding[29]. Therefore, subdivision of the population generates some degree of genetic differentiation between subpopulations[29]. Wright defined descriptive parameters that quantify the extent of the effects on heterozygosity due to subdivision, named F-statistics[30]. For two alleles at a locus it is possible to compute three F-statistics: the inbreeding coefficient – F_{IS} or F – which is a measure of the inbreeding within subpopulations and lies

between -1 and +1, with -1 meaning all individuals are heterozygous and +1 meaning all individuals are homozygous; the fixation index – F_{ST} – defined as a measure of the extent of genetic differentiation between subpopulations and varies between 0 and 1, with 0 meaning no differentiation and 1 meaning complete differentiation between subpopulations; and the overall fixation index – F_{IT} – which measures the reduction of heterozygosity relative to the total population.

The F-statistics were extended to multiple alleles by Nei[31], originating the G_{ST} parameter analog to F_{ST}, and then to multiple locus by averaging the values F_{ST} of all loci or by averaging the values of observed and expected heterozygosity over all loci, which is more appropriate. The value of G_{ST} for multiple alleles is equivalent to the weighted average of F_{ST} over all alleles[31]. Later on, Lynch and Milligan[32] developed a corrected method for computation of parameters to be applied to dominant molecular markers (like ISSRs), minimizing the generated bias due to the impossibility of distinction between homozygotes and heterozygotes.

Wright[33] also defined guidelines to interpret the F_{ST} values by creating four levels of genetic differentiation based on allozyme loci: little genetic differentiation, if F_{ST} is between 0 and 0.05; moderate genetic differentiation, if F_{ST} is between 0.05 and 0.15; great genetic differentiation, if F_{ST} is between 0.15 and 0.25; and very great genetic differentiation, if F_{ST} is greater than 0.25. Wright also noted that even if values of F_{ST} are below 0.05 they are not necessarily negligible. However, Hedrick (1999)[34] showed that in practical cases the maximum value of F_{ST} is not one but the expected level of homozygosity, meaning that extremely polymorphic molecular markers (like microsatellites) the F_{ST} values will always be low. Therefore, Wright's guidelines can give a general idea of the degree of genetic differentiation but should be evaluated with caution according to the chosen molecular marker.

1.4. Aims

In this study, we set out to compare the genetic diversity of *A. longifolia* in three different locations along the Portuguese coast – Vila Nova de Milfontes, Pinheiro da Cruz and Osso da Baleia – and determine if there is a correlation between the genetic diversity and the stark differences in phenology of this species found in the last two locations in a recent study. Furthermore, we also compare the genetic diversity of *A. longifolia* in these three locations with the genetic diversity of the deliberately introduced acacias in the coastline of Vila Nova de Milfontes in the late 1960s/early 1970s. Also, we conducted a more detailed study of the genetic diversity of *A. longifolia* in Vila Nova de Milfontes, assessing the genetic diversity of this species in this location alone.

With this study, we also intend to determine the origin of *A. longifolia* in Portugal, namely if the introduction of this species in Portugal occurred only through the plantation by the forestry services, and its pattern of dispersion throughout the Portuguese coastal line.

2. Materials and Methods

2.1. Plant Material

Ten phyllodes of *Acacia longifolia* were collected per accession and stored at -20°C until DNA extraction. Only green phyllodes were collected from individuals of approximately the same size. DNA extraction was performed as soon as possible, and after the extraction process the phyllodes were stored at -80°C. In Osso da Baleia (denoted OB) and Pinheiro da Cruz (denoted PC), 25 samples were collected in the vicinity of a set of central geographic coordinates (see Table 6.1 and Table 6.1 in Appendix I, respectively) in November and December 2015, respectively. In Vila Nova de Milfontes, samples were collected from different sites at this location for a more detailed study, and the geographic coordinates were determined for each individual (Table 6.3 in Appendix I). Samples from the 1960s/70s site (denoted MFAnt, 14 samples) and from the recently invaded site (denoted MFRec, 25 samples) were collected in September 2015 along with a sample of *Acacia saligna* to serve as an external control for the analysis, while samples from the coastal line (denoted MFCL, 14 samples) were collected in January 2016. The geographical distribution of the samples is shown in Figure 2.1.

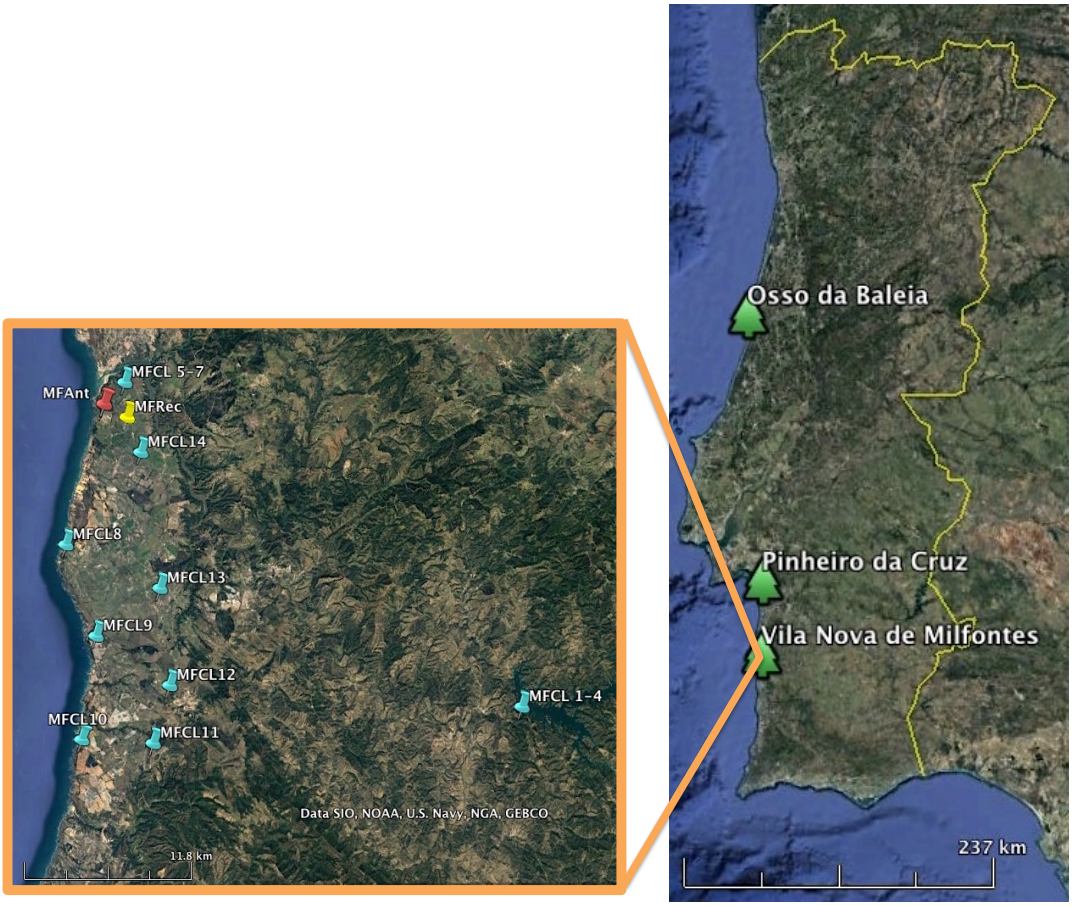

Figure 2.1: Locations of sample collection. Right: Map of Portugal containing the location of the samples from Osso da Baleia (OB), Pinheiro da Cruz (PC) and Vila Nova de Milfontes. Left: Map of Vila Nova de Milfontes containing the various locations of the samples. Red: MFAnt; Yellow: MFRec; Blue: MFCL. Pictures from Google Earth (accessed on September 5th 2016).

2.2. DNA Extraction and Quantification

Three different extraction protocols were tested with 3 samples of *Acacia longifolia* and the external control sample, *A. saligna*: the mini-CTAB method with ethanol; the mini-CTAB method with isopropanol, both adapted from Doyle & Doyle (1987)[35]; and the GF-1 plant DNA extraction kit (Vivantis, Malaysia). For detailed protocols see Appendix II. In all the protocols, *A. longifolia* phyllodes were macerated with liquid nitrogen and incubated at 65 °C with a 2% CTAB extraction buffer. After extraction, DNA was quantified by spectrophotometry, diluted to 10 µM and amplified through a PCR reaction with the ISSR primer 817 to confirm viability of the samples. The mini-CTAB method where DNA was precipitated with ethanol yielded the best results and was therefore chosen for this species.

2.3. Inter-Simple Sequence Repeats (ISSRs)

Each PCR reaction was performed in a total volume of 15 µL containing 6 ng of genomic DNA, 1 µM of each ISSR primer (STAB Vida, Portugal), 0.2 mM of dNTP mix, 1 mg/mL of BSA, 1.5 mM of $MgCl_2$, 0.6 U of GoTaq Flexi DNA polymerase and 1x green GoTaq Flexi buffer (Promega, USA). A total of 31 primers were tested but only a few were selected for further analysis: 16 were selected for analysis of genetic diversity of *A. longifolia* in the 3 collection sites and 11 were selected for a detailed study of this species' diversity in Vila Nova de Milfontes (see Table 6.4 in Appendix III). Primers that showed no amplification or low polymorphism level were excluded. Each PCR reaction included a negative control with all the components except the genomic DNA. The amplification was performed in a BioRad T100 thermocycler and included an initial denaturation step at 95°C for 5 min; followed by 40 cycles of 45 s at 95°C, 45 s at the corresponding annealing temperature for each primer (Table 6.4 in Appendix III), and 1.5 min at 72°C; followed by a final extension step at 72°C for 7 min. Primers 840, 846, 901, 903 and 904 required a touch-down amplification program as follows: initial denaturation step at 95°C for 5 min; followed by 11 cycles of 45 s at 95°C, 45 s at 57°C, and 1.5 min at 72°C; followed by 29 cycles of 45 s at 95°C, 45 s at 52°C (primers 901, 903 and 904), 53°C for primer 846 or 54°C for primer 840, and 1.5 min at 72°C; followed by a final extension step at 72°C for 7 min. The amplification products were separated by electrophoresis in a 2% agarose gel with 2 µL of GreenSafe per 100 mL of agarose, using 1x TAE as both the gel buffer and the running buffer. Gels were run at 60 V for 4h and were visualized in the GeneFlash UV transilluminator (Syngene, UK). Each gel image was analyzed with the GeneTools software (Syngene, UK) and only bands with sizes between 250 and 2.400 bp were considered in the analysis. For each primer, a binary matrix was filled with either a 1 or a 0, whether a certain band was present or absent, respectively. At least 10% of the PCR amplifications were performed twice to confirm the reproducibility of the technique. The

data from all the primers was then combined and computed using NTSYS v2.1[36] software to obtain the dendrograms by UPGMA with Dice's coefficient. Descriptive parameters were computed using the AFLP-Surv v1.0[37] software.

2.4. Simple Sequence Repeats (SSRs) or Microsatellites

Primers designed for other species of *Acacia* reported by Roberts *et al.*[27] and Forrest *et al.*[28] were selected to check for amplification in *A. longifolia*. A total of 11 pairs of microsatellite primers were tested and three were selected for further analysis (Table 6.5 in Appendix IV). Most primers showed amplification with this species but required some optimization, namely the annealing temperature of the PCR cycle, concentration of $MgCl_2$ or quantity of genomic DNA. The selected primers, amplifying locus APZIZ, DCLOC and CPUH4, showed mostly low molecular weight amplification products. The forward primers for these loci were labeled with ATTO-550, 6-FAM and HEX, respectively, for capillary electrophoresis. See Table 6.6 in Appendix IV for the selected primer's details. Each PCR reaction was performed in a total volume of 15 µL containing 10 ng of genomic DNA, 0.2 µM of forward primer labeled with a fluorescent dye on the 5' end (STAB Vida), 0.2 µM of reverse primer (STAB Vida), 0.2 mM of dNTP mix, 2 mM of $MgCl_2$, 1 U of GoTaq Flexi DNA polymerase and 1x colorless GoTaq Flexi buffer (Promega, USA). Each PCR reaction included a negative control containing all the components except the genomic DNA. Twelve repeated samples were performed for confirmation of reproducibility. Twelve positive controls (random selected samples) were also performed for confirmation of amplification before proceeding to capillary electrophoresis. The amplification was performed in a BioRad T100 thermocycler as follows: initial denaturation step at 95°C for 5 min; followed by 40 cycles of 30 s at 94°C, 1 min at 59°C, and 30 s at 72°C; followed by a final extension step at 72°C for 7 min. The positive and negative controls were analyzed by electrophoresis in a 2% agarose gel prepared as described before, at 80 V for 2h. The results were then visualized as previously described. Each PCR with a given pair of primers was performed separately and then the amplification products were combined for simultaneous analysis. The samples were analyzed through capillary electrophoresis by STAB Vida (Portugal). The results were analyzed with Peak Scanner v1.0 software (Applied Biosystems, USA) with GS500(-250) as size standard. Descriptive parameters and allelic frequencies were calculated using the GenAlEx v6.502[38,39] software.

3. Results and Discussion

3.1. Inter-Simple Sequence Repeats (ISSR) Analysis

3.1.1. *Acacia longifolia* through Portugal

A total of 31 ISSR primers were tested and 16 were selected to assess the genetic variability of *Acacia longifolia* in Vila Nova de Milfontes (MFAnt and MFRec), Osso da Baleia (OB) and Pinheiro da Cruz (PC). The choice of the primers was based on the amplification of clear and polymorphic bands as shown in the agarose gel in Figure 3.1. A total of 275 bands (243 excluding the external control) were obtained and 253 were polymorphic (Table 3.1), resulting in the dendrogram presented in Figure 3.2. The dendrogram shows a high genetic differentiation of all *A. longifolia* samples from the external control (EC) *A. saligna* (S_{Dice} = 0.20). Separation among *A. longifolia* samples occurred at a much higher Dice's coefficient (first separation at S_{Dice} = 0.63) and the clusters are all close together (last clustering at S_{Dice} = 0.85), indicating that these samples share a high degree of genetic similarity despite their geographical collection site.

Table 3.1: Total number of bands with and without the external control (EC), number and percentage of polymorphic bands (PB) for each ISSR primer for analysis of samples from Vila Nova de Milfontes (MFAnt, MFRec), Osso da Baleia (OB) and Pinheiro da Cruz (PC).

Primer	Nº Total Bands (with EC)	Nº Total Bands (without EC)	Nº PB (with EC)	%PB (with EC)
808	18	16	18	100%
810	19	17	19	100%
813	24	23	23	96%
817	14	11	11	79%
825	25	24	25	100%
826	11	6	9	82%
827	17	14	17	100%
836	20	19	19	95%
840	14	14	14	100%
846	15	14	15	100%
849	13	11	12	92%
868	15	13	14	93%
880	10	10	10	100%
901	25	23	23	92%
903	20	16	18	90%
904	15	12	11	73%
TOTAL	**275**	**243**	**258**	**94%**

Analyzing the dendrogram, the first separation originates two clusters: cluster I and cluster II. Cluster I then subdivides into groups A and B. Group A includes all samples from MFAnt and three

samples from MFRec. Group B subdivides into two subgroups: subgroup 1 contains most of the samples from OB; and subgroup 2 contains all samples from PC. Cluster II is divided into groups C and D: group C contains most of the samples from MFRec and group D contains nine samples from OB. Therefore, samples cluster according to their geographical location, which seems to be indicative of some differentiation.

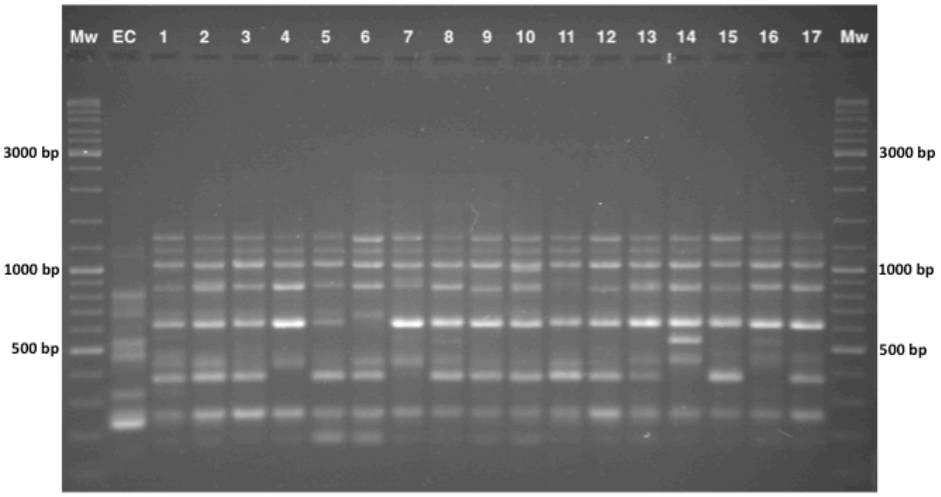

Figure 3.1: Agarose gel of the amplification products obtained with ISSR primer 849. Mw – molecular weight standard, with the three most intense bands (3.000, 1.000, 500 bp) indicated; EC – external control (*A. saligna*); Lanes 1-17 – samples of *A. longifolia*.

The computation of descriptive population genetic parameters allowed a better understanding of the genetic diversity of *A. longifolia* throughout Portugal (Table 3.2). The obtained Nei's gene diversity (*h*) values are low in general (mean *h* = 0.190, equivalent to H_S), with the highest value belonging to MFAnt (*h* = 0.213) and the lowest value belonging to PC (*h* = 0.166), indicating that acacias found in PC are genetically less diverse than acacias found in MFAnt (discussed later).

Table 3.2: Descriptive parameters of genetic variability of *A. longifolia* obtained with ISSR markers for samples from Vila Nova de Milfontes (MFAnt and MFRec), Osso da Baleia (OB) and Pinheiro da Cruz (PC). Nº PB – number of polymorphic bands without the external control; %PB – percentage of polymorphic bands; *h* – Nei's gene diversity, analogous to the expected heterozygosity under Hardy-Weinberg equilibrium; H_S – mean expected heterozygosity within subpopulations; H_T – expected heterozygosity in total population; F_{ST} – Wright's fixation index. Standard error is shown in parenthesis. *P(rand ≥ data) < 0.01 based on 500 permutations.

Location	Nº PB (without EC)	%PB	*h*	H_S	H_T	F_{ST}
MFAnt	135	55.6	0.213 (0.012)	0.190 (0.010)	0.204	0.070* (0.098)
MFRec	129	53.1	0.187 (0.011)			
OB	136	56.0	0.194 (0.011)			
PC	121	49.8	0.166 (0.011)			

The computation of the Wright's fixation index showed a moderate genetic diversity, with a value of 0.070 (P(rand ≥ data) < 0.01). This supports the hypothesis stated above that samples of *A. longifolia* from the four locations under analysis have a relatively high level of similarity at a molecular level. However, there is low but significant genetic differentiation between locations. Furthermore, to check the values of the fixation index between each and every location, the pairwise F_{ST} matrix was computed (Table 3.3).

Table 3.3: Pairwise F_{ST} matrix for samples from Vila Nova de Milfontes (MFAnt and MFRec), Osso da Baleia (OB) and Pinheiro da Cruz (PC) obtained from ISSR markers.

	MFAnt	MFRec	OB	PC
MFAnt	0.000			
MFRec	0.056	0.000		
OB	0.072	0.046	0.000	
PC	0.094	0.078	0.074	0.000

The lowest value of F_{ST} was obtained between MFRec and OB (F_{ST} = 0.046), indicating that samples in these two locations were the most similar. This is supported by the fact that MFRec and 9 samples from OB clustered together on the dendrogram (Cluster II). The highest F_{ST} value was obtained between MFAnt and PC (F_{ST} = 0.094), indicating that these two locations were the most genetically different and indeed these two regions are located in different groups (A and B) in Cluster I. Also important to notice is the F_{ST} between MFAnt and MFRec (F_{ST} = 0.056), which is the second lowest value of the matrix. This seems to indicate some kind of genetic differentiation between these two locations, which belong to different clusters in the dendrogram, even though they are genetically similar and geographically close. Also, the fact that OB shows different F_{ST} values with both MFAnt (F_{ST} = 0.072) and MFRec (F_{ST} = 0.046) is indicative of some genetic differentiation between these last two locations, with OB being more similar to MFRec than the MFAnt location. Still, it is important to notice that three samples from MFRec appeared together with the samples from MFAnt in the dendrogram, which supports the fact that these locations also have some degree of similarity despite their separation in different clusters. Also, the highest F_{ST} values computed belong all to PC, indicating that this was the most differentiated location of all four. In short, all F_{ST} values but one fall into Wright's moderate genetic differentiation category and seem to vary between locations, evidencing low but existent genetic differentiation that might explain why the samples have a tendency of grouping by their geographical site. The exception is the F_{ST} value between MFRec and OB, which falls into Wright's low genetic differentiation category. However, the F_{ST} values are low in general, reinforcing the hypothesis that all locations had a considerable degree of genetic similarity despite their geographical distance. This might indicate a single origin of *A.*

longifolia in Portugal, and that the stark differences in this species' phenology found in Osso da Baleia and Pinheiro da Cruz by Fernandes *et al.* (2015)[8] are consequence of its amazing phenotypic plasticity and not of genotypic variation.

A similar approach to the one presented in this study reports the genetic diversity of *Oeceoclades maculata*, an orchid original from Africa but invasive in Brazil[40]. The authors collected samples from several sites in Brazil and performed ISSR analysis based on 13 primers, yielding a total of 192 bands, 189 of which were polymorphic. The Nei's gene diversity (*h*) obtained for *O. maculate* is 0.2556, higher than the one obtained in this study for *A. longifolia* (*h* = 0.190) and indicative of a greater genetic diversity of *O. maculate* when compared to *A. longifolia*. The authors also computed the Φ_{ST} parameter, which is a measure of genetic differentiation, and obtained a value of 0.933, indicating great genetic differentiation between the Brazilian populations of *O. maculate*, unlike the case of *A. longifolia* in Portugal. Another study, this time with AFLP markers, compared native and invasive populations of *Solidago canadensis*, a species original from North America and introduced in China[41]. The authors computed the Φ_{PT} parameter, an analog of F_{ST}, and found that the North American populations showed lower genetic differentiation than the Chinese populations (Φ_{PT} = 0.103 vs. Φ_{PT} = 0.182, respectively), which might be indicative, among other possibilities, of multiple native populations as the origin of this species in China. Conversely, the results obtained for *A. longifolia* show low genetic differentiation (F_{ST} = 0.070) and seem to be indicative of a single origin of this species in Portugal.

Regarding the genus *Acacia*, not much has been done regarding molecular analysis. *A. Senegal,* a native species from four districts of Kenya[21], was the subject of a study where 5 ISSR markers and 10 RAPD markers were combined. The mean Nei's gene diversity value obtained was 0.283, which was higher than the one obtained with *A. longifolia* (*h* = 0.190), indicating that *A. senegal* has a higher level of genetic diversity in Kenya than *A. longifolia* has in Portugal. This is to be expected since *A. senegal* is native in Kenya, but *A. longifolia* is an invader introduced in Portugal. Still, evidence was found that there is some genetic differentiation between populations of *A. senegal* in Kenya, and the authors raise the possibility that it might be consequence of geographical separation of the populations. Conversely, *Vachellia (Acacia) karroo*, a native species from South Africa, was also subject to a study using ISSR markers, and it was found that this species is a panmictic entity[42], which means that there are no subpopulations genetically differentiated even if geographically distanced. These are some of the few studies reported in the genus Acacia using ISSRs, although these molecular markers have been used in many diversity studies with numerous other plant species, including *Eucalyptus grandis*[43] and an invasive species from Portugal: *Pittosporum undulatum*[20].

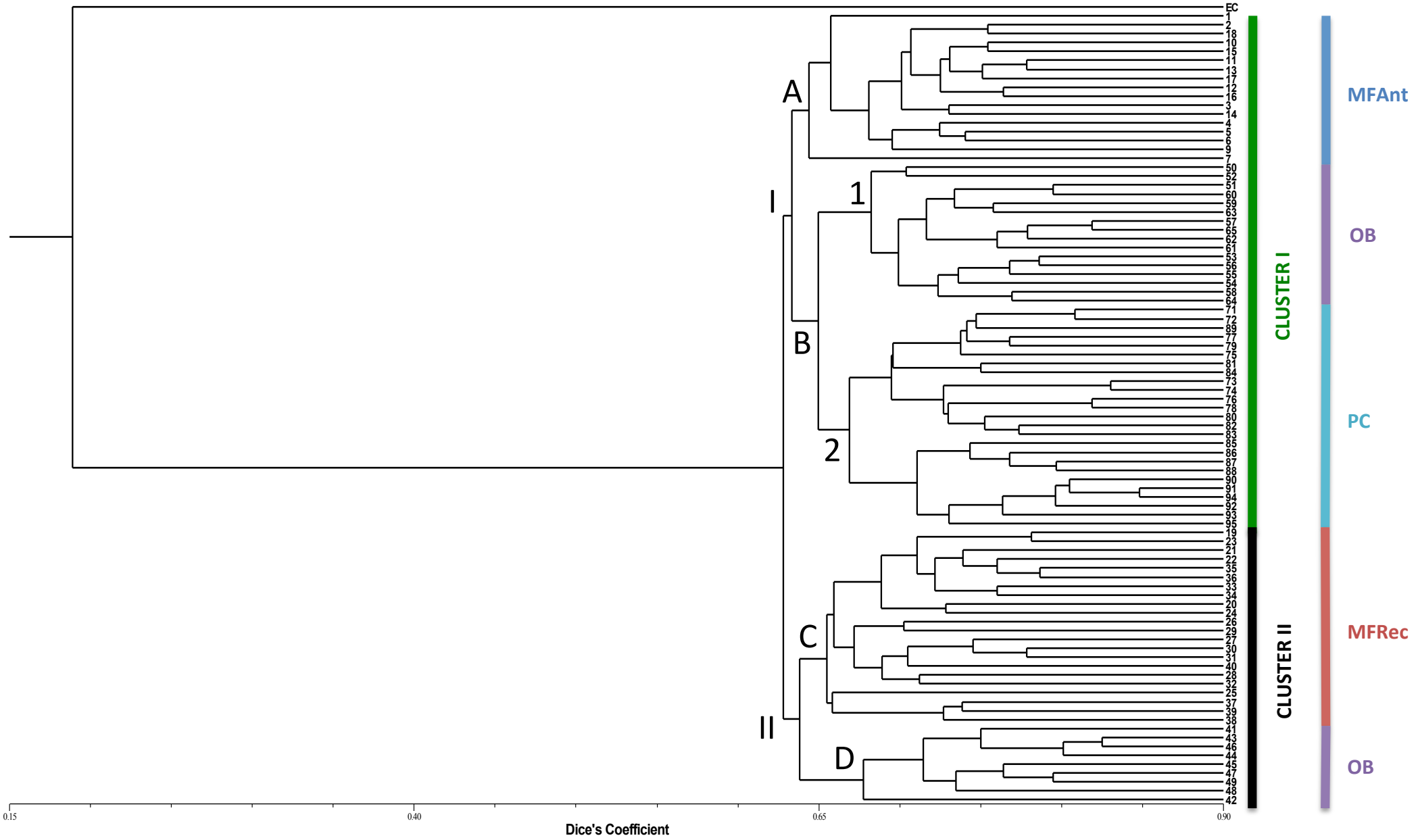

Figure 3.2: Dendrogram obtained with ISSR markers with samples from Vila Nova de Milfontes (MFAnt and MFRec), Osso da Baleia (OB) and Pinheiro da Cruz (PC). EC – external control; Samples 1-15 – MFAnt; Samples 16-40 – MFRec; Samples 41-65 – OB; Samples 71-95 – PC.

3.1.2. *Acacia longifolia* in Vila Nova de Milfontes

For the detailed analysis of Vila Nova de Milfontes only 11 primers revealing high polymorphism were selected (see table 4 in Appendix III), generating a total of 137 bands (128 excluding the external control), all of which are plolymorphic (Table 3.4). The dendrogram presented in Figure 3.3 shows again a clear separation of the external control *A. saligna* from all the *A. longifolia* samples (S_{Dice} = 0.31). The samples from Vila Nova de Milfontes shared roughly the same similarity level obtained with all the sampling (first level at S_{Dice} = 0.61, last level at S_{Dice} = 0.79 *vs* S_{Dice} = 0.63, last level S_{Dice} = 0.85 as previously reported). However, neither MFAnt nor MFRec or MFCL samples clustered according to the collection sites.

Table 3.4: Total number of bands with and without the external control (EC), number and percentage of polymorphic bands (PB) for each ISSR primer for analysis of samples from Vila Nova de Milfontes (MFAnt, MFRec and MFCL).

Primer	Nº Total Bands (with EC)	Nº Total Bands (without EC)	Nº PB (with EC)	%PB (with EC)
808	11	11	11	100%
810	17	15	17	100%
813	10	9	10	100%
823	13	13	13	100%
827	11	8	11	100%
836	11	10	11	100%
840	14	14	14	100%
846	10	10	10	100%
849	14	14	14	100%
868	9	8	9	100%
901	17	16	17	100%
TOTAL	**137**	**128**	**137**	**100%**

The descriptive parameters were also computed for this data as previously (Table 3.5). The values of Nei's gene diversity (*h*) were low in general (H_S = 0.261) and quite similar among sites, with the highest value again belonging to MFAnt (*h* = 0.275) and the lowest belonging to MFRec (*h* = 0.246). The obtained value of F_{ST} was below 0.05 (F_{ST} = 0.013, P(rand ≥ data) < 0.01), corresponding to a Wright's low genetic differentiation category, indicating high genetic similarity between sites. However, this value was significant and, as mentioned before, not necessarily negligible, but it was much lower than the one obtained previously when considering Osso da Baleia, Pinheiro da Cruz and Vila Nova de Milfontes all together (F_{ST} = 0.070).

Table 3.5: Descriptive parameters of genetic variability of *A. longifolia* obtained with ISSR markers for samples from Vila Nova de Milfontes (MFAnt, MFRec and MFCL). Nº PB – number of polymorphic bands without the external control; %PB – percentage of polymorphic bands; *h* – Nei's gene diversity, analogous to the expected heterozygosity under Hardy-Weinberg equilibrium; H_S – mean expected heterozygosity within subpopulations; H_T – expected heterozygosity in total population; F_{ST} – Wright's fixation index. Standard error is shown in parenthesis. *P(rand ≥ data) < 0.01 based on 500 permutations.

Location	Nº PB (without EC)	%PB	*h*	H_S	H_T	F_{ST}
MFAnt	86	67.2	0.275 (0.016)	0.261 (0.009)	0.265	0.013* (0.046)
MFRec	86	67.2	0.246 (0.015)			
MFCL	89	69.5	0.263 (0.016)			

The pairwise F_{ST} matrix was computed (Table 3.6), revealing in general low values, not very different from one another. Also, they were much lower than those obtained previously with samples from Vila Nova de Milfontes (MFAnt and MFRec only), Osso da Baleia and Pinheiro da Cruz. The lowest F_{ST} value (0.009) was obtained between MFAnt and MFRec, indicating that these sites were the most similar. Interestingly, the F_{ST} value previously obtained between these two sites (F_{ST} = 0.056) was much higher, suggesting some sort of genetic differentiation. We believe that this was caused by the presence of data from OB when computing the F_{ST}. When considering the MFAnt, MFRec and OB locations, MFRec was indeed slightly more similar to OB than to MFAnt. However, when comparing MFAnt and MFRec with MFCL, the first two sites seemed to be much more similar than before. This indicates that MFAnt and MFRec despite being most similar clearly have low but existing genetic differentiation. This might justify why MFAnt and MFRec sites appeared in different clusters of the dendrogram obtained with samples from all locations (Figure 3.2) despite their geographical proximity. The highest F_{ST} (0.017) belonged to MFRec and MFCL, indicating that these two sites were the most genetically different of the three.

Table 3.6: Pairwise F_{ST} matrix for samples from Vila Nova de Milfontes (MFAnt, MFRec and MFCL) obtained from ISSR analysis.

	MFAnt	MFRec	MFCL
MFAnt	0.000		
MFRec	0.009	0.000	
MFCL	0.013	0.017	0.000

However, as previously mentioned, all F_{ST} values were very low and seemed to indicate great genetic similarity of *A. longifolia* across Vila Nova de Milfontes, raising the hypothesis that the individuals from the MFRec and MFCL sites originated from the MFAnt site that spread and invaded the nearby soils.

Comparing the obtained results with the previously mentioned studies of *Oeceoclades maculata*[40] and *Solidago canadensis*[41], it is clear that the genetic differentiation of *A. longifolia* found in the three sites of Vila Nova de Milfontes is much lower, which is to be expected considering that since these studies included samples from a higher area when compared to our Vila Nova de Milfonte's study area, and also considering that the genetic differentiation between Vila Nova de Milfontes, Osso da Baleia and Pinheiro da Cruz is not high even considering the distance that separates these locations. However, the mean value of Nei's gene diversity obtained for Vila Nova de Milfontes alone is higher than the one obtained previously for all three locations (H_S = 0.190), indicating that the diversity of acacias from Vila Nova de Milfontes is higher than the average along the country.

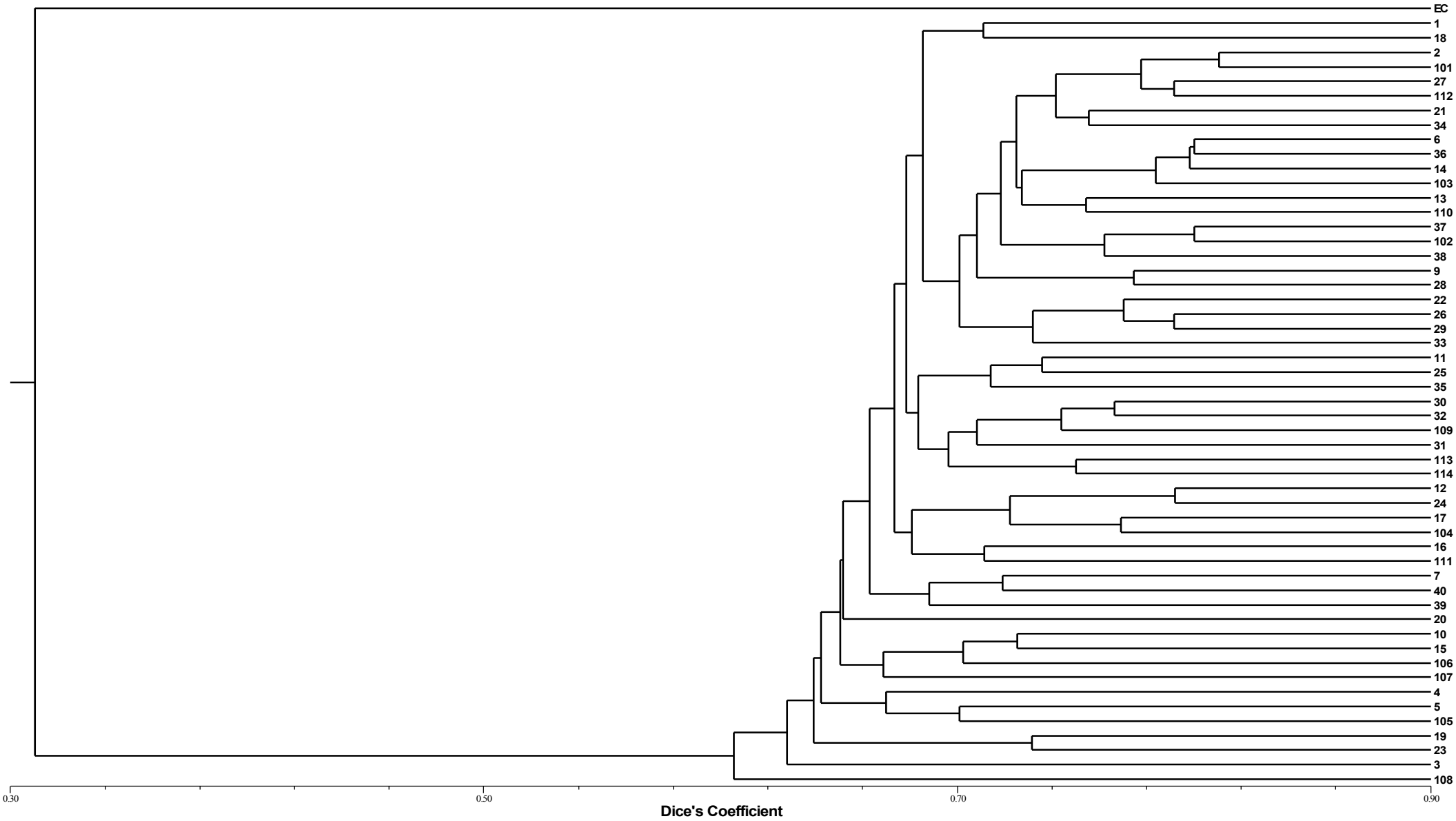

Figure 3.3: Dendrogram obtained with ISSR markers of samples from Vila Nova de Milfontes (MFAnt, MFRec and MFCL). EC – external control; Samples 1-15 – MFAnt; Samples 16-40 – MFRec; Samples 101-114 – MFCL (samples collected along the coastal line of Vila Nova de Milfontes).

3.2. Simple-Sequence Repeats (SSRs) or Microsatellites Analysis

Eleven pairs of microsatellite primers were tested in *A. longifolia* and three pairs were selected for the final analysis - the pairs of primers that amplified the locus APZIZ, DCLOC and CPUH4 according to Roberts *et al.*[27] and Forrest *et al.*[28]. However, after capillary electrophoresis the results with the primers APZIZ were not considered due to the unspecific amplification of high molecular weight fragments (visible as several peaks occurring in some samples). Therefore, they were excluded from the final analysis. The results with the pair of primers CPUH4 led to the exclusion of samples 9, 26, 29, 61, 79 and 91 from the final analysis. The reason for excluding samples 9 and 29 was the absence of amplification, while the remaining mentioned samples showed more than two peaks in the electropherogram, indicating that some other region of the genome was being amplified together with the microsatellite. This problem also happened with samples 1, 12, 13, 14, 25, 37, 46, 62, 78, 86 and 90 but in these cases we were able to determine which peaks resulted from microsatellite amplification and which peaks had to be excluded.

Analyzing the resulting electropherograms obtained for locus DCLOC and CPUH4 resulted in a total of 7 alleles: alleles 98, 102 and 108 bp for DCLOC; and alleles 93, 95, 98 and 121 bp for CPUH4. No private alleles were found within these loci.

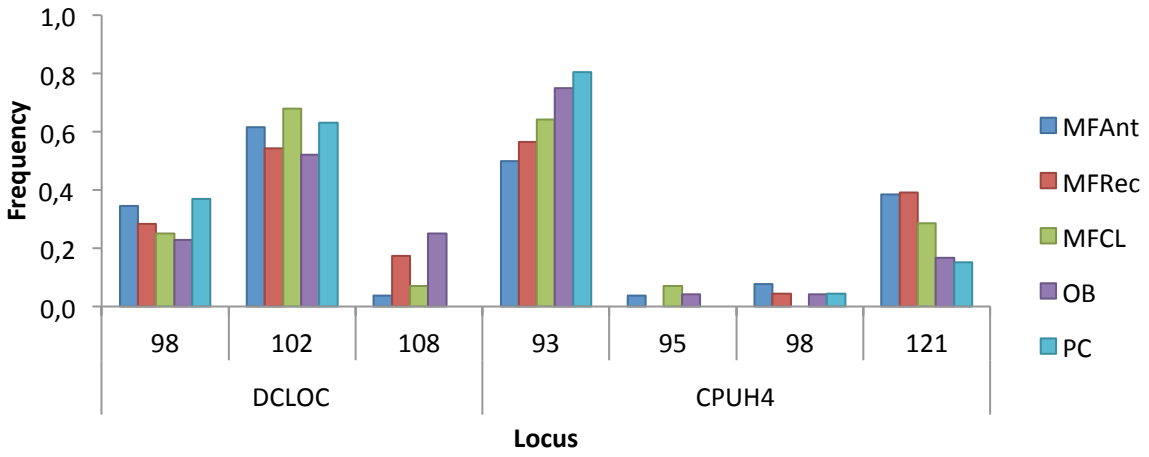

Figure 3.4: Allelic frequencies of loci DCLOC and CPUH4 for samples from Vila Nova de Milfontes (MFAnt, MFRec and MFCL), Osso da Baleia (OB) and Pinheiro da Cruz (PC).

Locus DCLOC has two alleles with high frequency in all populations – alleles 98 and 102 – while allele 108 has a lower frequency and was absent from the PC location (Figure 3.4). In locus CPUH4, two alleles were present at higher frequency – alleles 93 and 121 – and two alleles showed lower frequency – alleles 95 and 98. Allele 95 was absent from the MFRec and PC locations while

allele 98 was absent from the MFCL site. The descriptive statistics for each location are presented in Table 3.7.

Table 3.7: Descriptive statistics for samples from Vila Nova de Milfontes (MFAnt and MFRec), Osso da Baleia (OB) and Pinheiro da Cruz (PC) obtained with microsatellites. SE – standard error; *N* – number of samples. *A* – number of alleles detected; *Ae* – number of effective alleles; *I* – Shannon's information index; H_o – observed heterozygosity; H_e – expected heterozygosity; F_{IS} – Inbreeding coefficient; HWE – *P-value* for Hardy-Weinberg Equilibrium Chi-square test; ns – non-significant; H_T – mean expected heterozygosity in total population over all loci; F_{ST} – Wright's fixation index. *P(rand ≥ data) > 0.05 based on 999 permutations. No standard error computed for F_{ST} because of insufficient data (minimum 5 loci).

Location		*N*	*A*	A_e	*I*	H_o	H_e	F_{IS}	HWE (*P-value*)	H_T	F_{ST}
MFAnt	**Mean**	13	3.5	2.234	0.914	0.500	0.547	0.086	ns	0.520	0.038*
	SE	---	0.5	0.234	0.123	0.038	0.047	0.009			
MFRec	**Mean**	23	3.0	2.287	0.909	0.543	0.560	0.023	ns		
	SE	---	0.0	0.179	0.083	0.022	0.034	0.099			
MFCL	**Mean**	14	3.0	1.947	0.814	0.429	0.486	0.109	< 0.01 (locus CPUH4)		
	SE	---	0.0	0.053	0.016	0.143	0.014	0.320			
OB	**Mean**	24	3.5	2.136	0.902	0.625	0.510	-0.209	ns		
	SE	---	0.5	0.452	0.122	0.167	0.104	0.081			
PC	**Mean**	23	2.5	1.680	0.628	0.478	0.397	-0.203	ns		
	SE	---	0.5	0.192	0.030	0.087	0.069	0.010			
	Average	19.4	3.1	2.057	0.834	0.515	0.500	-0.039			
	SE	1.614	0.180	0.113	0.047	0.042	0.028	0.069			

All five populations were on Hardy-Weinberg equilibrium with the exception of the MFCL, which failed the chi-square HW equilibrium test for locus CPUH4 (*P-value* < 0.01). The lowest and most different value of Shannon's information index (*I*) was obtained with the PC location, indicating some genetic differentiation. The three highest Shannon's information index values were very similar and belonged to MFAnt, MFRec and OB, suggesting similarity among the three. The F_{IS} values show that the OB and PC locations had an excess of heterozygotes, while all the other locations had slightly more homozygotes than heterozygotes. Overall, there is slight excess in heterozygotes (average F_{IS} = - 0.039), which is indicative of low genetic diversity.

Due to the low number of loci analyzed and also the problems encountered with locus CPUH4 that led to the exclusion of individuals from this analysis, all parameters calculated with these data should be interpreted with caution. Still, we proceeded to the computation of the F_{ST} value and the pairwise F_{ST} matrix, similar to what was done for the ISSR data. The F_{ST} value obtained for our data, 0.038, was found to be non-significant (*P-value* > 0.05). This was probably due to insufficient data to support evidence of a true genetic differentiation between locations, even if low. Still, with some reservations, we can compare these results with those obtained with ISSR markers to see if there are any consistent patterns that would possibly corroborate any of the proposed hypotheses.

Table 3.8: Pairwise F_{ST} matrix for samples from Vila Nova de Milfontes (MFAnt and MFRec), Osso da Baleia (OB) and Pinheiro da Cruz (PC) obtained from microsatellite analysis.

	MFAnt	MFRec	MFCL	OB	PC
MFAnt	0.000				
MFRec	0.008	0.000			
MFCL	0.012	0.013	0.000		
OB	0.041	0.022	0.021	0.000	
PC	0.039	0.040	0.020	0.027	0.000

The lower F_{ST} values were obtained among MFAnt, MFRec and MFCL collection sites (Table 3.8), which support the ISSR data indicating that these locations were similar and that MFRec and MFCL were originated by invasion from the MFAnt location. Also as seen before with the ISSR data, the PC location had relatively high values of F_{ST} with all other locations when considering microsatellites, and the second lowest value was obtained with OB, supporting the clustering of these two locations in group B of Cluster I in the dendrogram (Figure 3.2). Interestingly, the F_{ST} value between MFAnt and OB was the highest obtained, and a relatively high F_{ST} value between these two locations was also obtained with the ISSR markers (see Table 3.3). Furthermore, in a similar way as with the ISSRs, the OB location has a lower F_{ST} value with MFRec than with MFAnt, supporting the clustering of OB and MFRec in Cluster II of the dendrogram (Figure 3.2) and the hypothesis that MFRec and MFAnt have some degree of genetic differentiation despite being the most similar. In general, and once again as obtained with the ISSR data, the F_{ST} values were low, supporting the hypothesis that the samples of *A. longifolia* from all five locations show some molecular differentiation. The presented microsatellite data supports the results obtained with the ISSR markers, which was a robust study based on a high number of polymorphisms. However, this microsatellite study should be regarded as preliminary and needs to be extended using a higher number of markers.

Microsatellites are an important and useful tool in molecular biology and have been used in studies of genetic variability of several invasive species. An example is *Ambrosia artemisiifolia*, a native species from North America but invasive in Pannonian Plain in Europe (includes Hungary, Serbia, Romania, Bosnia and Herzegovina and Croatia)[44]. This study included 5 microsatellite loci and yielded and F_{ST} value of 0.024, which is indicative of low genetic differentiation between populations. This F_{ST} value is lower than the value obtained for *A. longifolia* (0.038), but the later was found to be non-significant. Another example is species *Impatiens glandulifera*, which is native from India and Pakistan but invasive in other parts of the world including Canada, Finland and United Kingdom. In this study, the authors analyzed 10 microsatellite loci and compared the genetic differentiation between native and invasive populations and found that the native populations showed higher genetic differentiation than the invasive populations (F_{ST} = 0.216 vs. F_{ST} = 0.137,

respectively). They also performed sequencing of the nuclear ITS (Internal Transcribed Spacer) sequence and found evidence that *I. glandulifera* might have dispersed from Pakistan to Finland and Canada through the United Kingdom.

When one considers the genus *Acacia* the microsatellite methodology has been used widely in studies of genetic variability. Eight populations of native *A. senegal* from Uganda were studied by Mulumba *et al.*[45] based on four microsatellite loci, originating 8-14 alleles. The computed values of mean expected heterozygosity we obtained with *A. longifolia* (H_e = 0.500) were similar to those presented by the authors (H_e = 0.479). Also the mean Shannon's information index was similar between *A. senegal* in Kenya and *A. longifolia*, (0.927 and 0.834, respectively). However, *A. senegal* had a low but significant genetic differentiation among populations (F_{ST} = 0.100, P(rand ≥ data) < 0.001), unlike *A. longifolia*. Still, it is important to notice that *A. Senegal* is native in Uganda while *A. longifolia* is an introduced species in Portugal, which might have an effect on the genetic variability. Furthermore, the genetic variability of *A. senegal* was also studied in Kenya with 7 nuclear microsatellite loci plus two chloroplastidial loci[46]. Another introduced *Acacia* species in Portugal, *Acacia saligna*, was also subject of a study that intended to find genetic patterns of introduced populations using microsatellites and the ETS (External Transcribed Spacer) region[47]. Results show that there are no patterns of *A. saligna*, that is, all subspecies are found around the world, but a specific South African lineage also appeared in both Italy and Portugal.

Considering the limitations of the low amount of SSR data, we opted by performing a safer, simplistic analysis by classifying alleles as common (C) or rare (R) for each locus. We defined a rare allele as one that has the lowest general frequency and doesn't appear in all populations. Therefore, allele 108 from locus DCLOC and alleles 95 and 98 from locus CPUH4 were classified as rare (see Figure 3.4). All other alleles were considered as common. Using these guidelines, we genotyped all individuals with the CR system for locus DCLOC and CPUH4. For example, if a sample has a rare allele in locus DCLOC and CPUH4, it was genotyped as RR; if a sample has a common allele in locus DCLOC and a rare allele in locus CPUH4, it was classified as CR. All samples were genotyped as CC, CR, RC or RR. No distinction was established among samples that had the CPUH4 95 allele from those that had the CPUH4 98 allele. After genotyping, the frequency of the rare alleles in each location was checked, combining the data from individuals classified as CR, RC or RR (Figure 3.5).

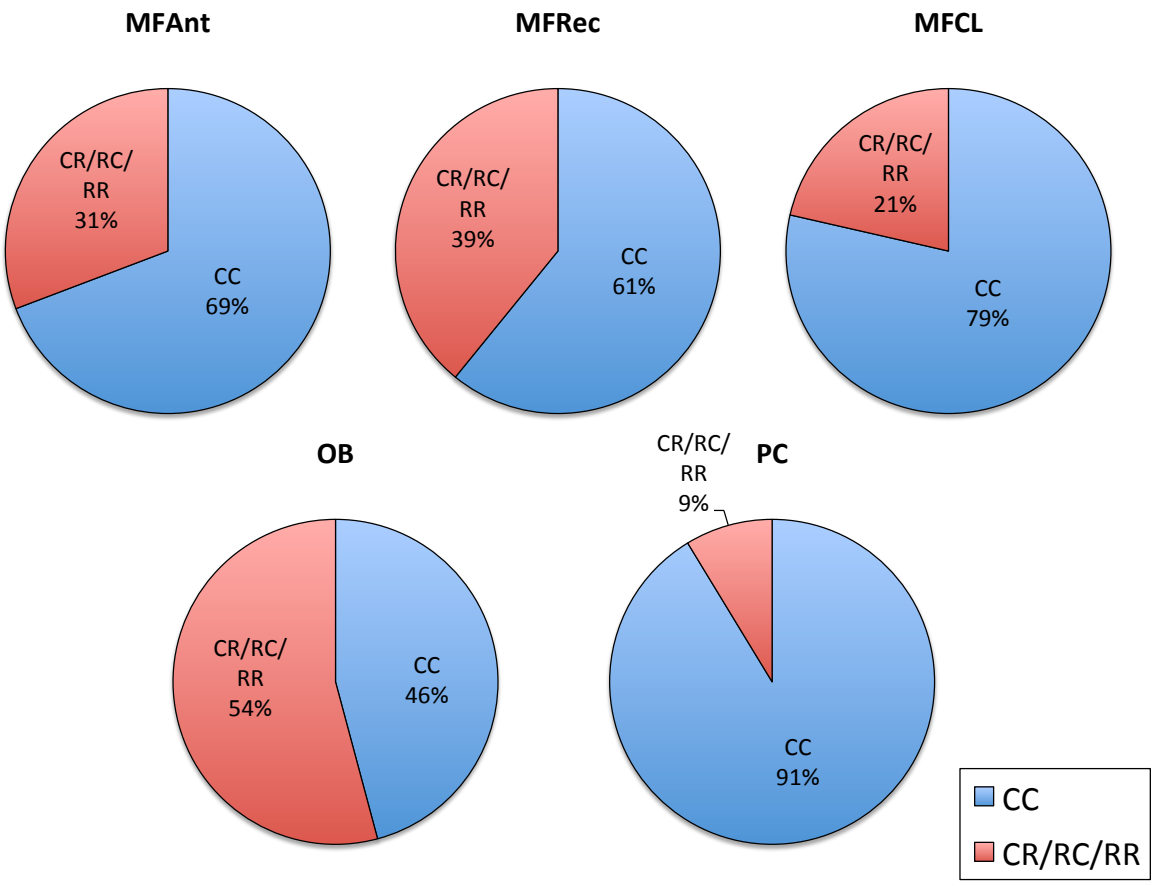

Figure 3.5: Frequency of the common (C) and rare (R) allelles in the *A. longifolia* samples from Vila Nova de Milfontes (MFAnt, MFRec and MFCL), Osso da Baleia (OB) and Pinheiro da Cruz (PC).

The MFAnt population had at least one rare allele in 31% of the samples. The geographically closest sites to MFAnt – MFRec and MFCL – had at least one rare allele in 39% and 21% of the samples, respectively. This supports the hypothesis that the MFRec and MFCL individuals originated from the MFAnt location since they share the same rare alleles and are genetically similar (also corroborated by ISSR analysis), which is in accordance with the geographical proximity. Interestingly, 54% of samples from OB also had at least one rare allele, which supports the genetic similarity obtained between this location and MFRec with ISSRs. This might also be a possible explanation as to why MFRec is both similar and slightly differentiated from MFAnt. These two sites shared the same rare alleles but in different percentages, higher in MFRec than in MFAnt. This might explain why MFRec is similar to both OB and MFAnt, despite the fact that MFAnt and the OB locations were slightly more genetically differentiated. This hypothesis needs confirmation, given the low number of locus analyzed. New microsatellite data will allow finding other private and shared alleles between these locations.

3.3. *Acacia longifolia* in Portugal: Origin and Dispersion Patterns

Considering both analysis with ISSRs and microsatellites, we showed that samples of *A. longifolia* from different locations in Portugal have low but significant genetic differentiation. However, samples from all location studied shared rare alleles, indicating a single, common origin for *A. longifolia* in Portugal. It is documented in forestry services records that this species was introduced in the Portuguese coastal line with the purpose of preventing the erosion of sand dunes and to protect the agriculture fields and forests in the vicinity[5–7]. In face of our results, our hypothesis is that acacias planted in the coastal line spread out and invaded adjacent soils. This hypothesis is supported by the fact that samples from MFRec and MFCL have high molecular similarity according to ISSR analysis and also share rare alleles with MFAnt. In other words, acacias that appeared more recently in Vila Nova de Milfontes are very similar to the ones that were introduced by the forestry services in the late 1960s/early 1970s, which supports the hypothesis that the recent acacias originated from the introduced ones. Regarding Osso da Baleia, we did not have access to historical records of plantation of *A. longifolia*. However, there are records of plantation of acacias in Quiaios-Mira sand dunes, which are located roughly 25 Km north. It is possible that (A) these acacias spread out towards Osso da Baleia, but we cannot exclude the possibility that (B) acacias were introduced in Osso da Baleia by the forestry services and spread to adjacent soils (see Figure 3.6 for visual representation). Interestingly, samples from this location showed similarity to samples from Vila Nova de Milfontes and also share the same rare alleles in relatively high frequency. Therefore, we hypothesize that, if indeed acacias from Osso da Baleia originated from the Quiaios-Mira sand dunes, then the acacias from the later location should be similar to those from MFAnt. This hypothesis seems probable if we consider that the forestry services might have used the same seed allotment for plantation on both sites, or if seeds were collected in Quiaios-Mira and planted in Vila Nova de Milfontes.

Our results also indicate that the samples from Pinheiro da Cruz are the most differentiated, and indeed they have the lowest frequency of rare alleles (9%). It is important to notice that, of all three locations, the acacias in Pinheiro da Cruz are the most isolated, and indeed the seeds from this location are low in number and have poor success when compared with Osso da Baleia[8]. The collection sites in Vila Nova de Milfontes and Osso da Baleia have public roads nearby with some traffic, which might help the invasive process. Pinheiro da Cruz, on the other hand, is an isolated location and the site is only assessable through a private road with restricted assess belonging to the "Estabelecimento Prisional de Pinheiro da Cruz". This may have caused isolation of the acacias contributing to a decrease in the rare alleles' frequency. So how did *A. longifolia* get to Pinheiro da Cruz according to our hypothesis? Either (1) a genetically similar group of acacias was deliberately

introduced in Pinheiro da Cruz by the forestry services, spreading and invading the adjacent soils, or (2) acacias from Vila Nova de Milfontes spread towards north and eventually reached Pinheiro da Cruz, losing rare alleles through the invasive process. It is also possible to consider that (3) acacias introduced in Costa da Caparica (Almada) by the forestry services spread south and reached Pinheiro da Cruz (see Figure 3.6 for visual representation). We also did not have access to any historical records of Pinheiro da Cruz, and both Vila Nova de Milfontes and Costa da Caparica are located about 60 Km away from this location. Therefore, all hypotheses seem equally plausible, even though hypothesis (3) would mean that acacias introduced in Costa da Caparica are also similar to acacias from MFAnt.

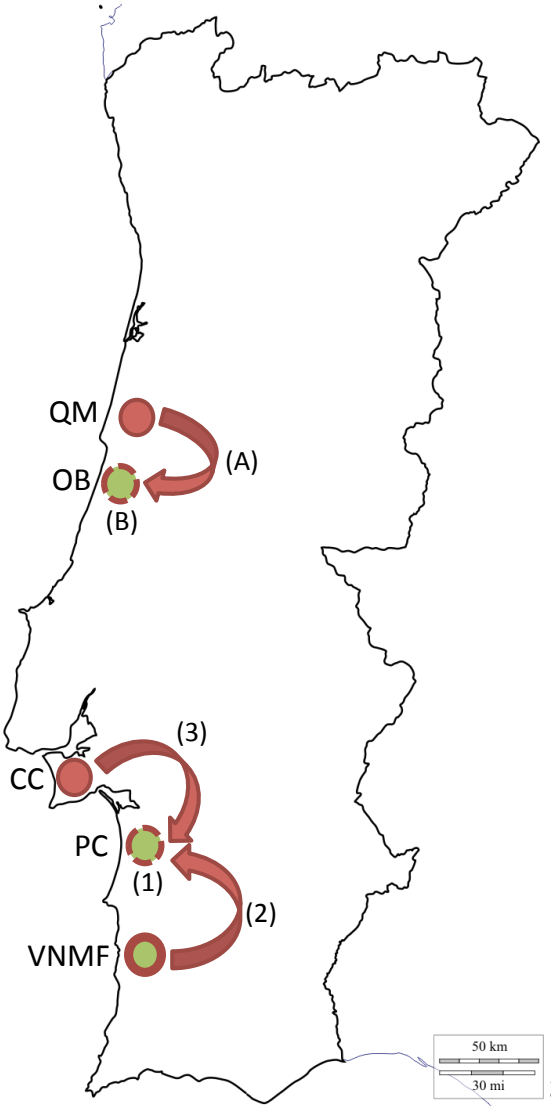

Figure 3.6: Representation of the hypotheses of distribution of *A. longifolia* in Portugal. Green – Studied locations; Red – Locations where *A. longifolia* was introduced by the forestry services; QM – Quiaios-Mira dunar system; OB – Osso da Baleia; CC – Costa da Caparica; PC – Pinheiro da Cruz; VNMF – Vila Nova de Milfontes; (A), (B) – Hypothesis of the origin of *A. longifolia* in Osso da Baleia; (1), (2), (3) – Hypothesis of the origin of *A. longifolia* in Pinheiro da Cruz.

It is important to notice, however, that even though low, we did find evidence of significant genetic differentiation between acacias from Osso da Baleia, Pinheiro da Cruz and Vila Nova de Milfontes. This low genetic differentiation might be the consequence of human intervention and agricultural management functioning as a driver of genetic variability, and might possibly lead to the formation of genetically distinct populations in the future.

As was previously mentioned, our microsatellite analysis is based on preliminary data, meaning that our results need confirmation in the future. Still, some results obtained with microsatellites are in accordance and corroborate the ISSR data, which brings some reassurance to our analysis. Further molecular analysis of *Acacia longifolia* has to be performed in the future, either through more microsatellites or other molecular markers. This will bring a clearer picture of the genetic variability and invasive route of this species in Portugal.

4. Conclusions and Future Perspectives

Samples of *A. longifolia* from the Portuguese coast showed low but significant genetic differentiation among locations, which seems to be indicative of a single introduction of this species in Portugal. This is supported by the fact that samples from all three locations studied share rare alleles. Also, the study of Vila Nova de Milfontes in particular evidenced much lower genetic differentiation between collection sites, indicating that acacias planted by the forestry services spread and invaded adjacent soils. Genetic similarity of acacias in Portugal can be due to sowing the same seed allotment or by transference of seeds from one location to another, and these germinated and then spread through the costal line. Therefore, we hypothesize that the stark differences in phenology found between acacias from Osso da Baleia and Pinheiro da Cruz are due to the amazing phenotypic plasticity of *A. longifolia* and not the result of genetic variation to the environment.

It is important to point that our microsatellite data has limitations, given the low number of loci analyzed. This resulted in the lack of significant evidence of genetic variability among samples from different collection sites, which, even if low, was shown to exist through our very robust ISSR analysis. Still, with a very simplistic and rough analysis of allele frequencies we were able to find some evidences that supported our ISSR data, which brings some confidence to our microsatellite results. However, more detailed studies with a higher number of microsatellite markers and even other types of molecular markers are needed to better understand the distribution of *A. longifolia* in Portugal.

Acacia longifolia is and aggressively invasive species that, even in the relatively short period of time since its introduction in the Portuguese coastal region, was able to spread throughout the country and is now found virtually everywhere in Portugal. It would be extremely important to find ways to prevent or slow down this species' invasion whether throughout mechanical or chemical ways. Alternatively, finding economically valuable usages of *A. longifolia* products, obtained either directly or indirectly, would also be a sustainable solution to control its invasion and could be beneficial in the long term.

5. References

[1] J. D. de Almeida and H. Freitas, "The exotic and invasive flora of Portugal.," *Bot. Complut.*, vol. 25, pp. 317–327, 2001.

[2] S. Rodríguez-Echeverría, J. A. Crisóstomo, C. Nabais, and H. Freitas, "Belowground mutualists and the invasive ability of Acacia longifolia in coastal dunes of Portugal," *Biol. Invasions*, vol. 11, pp. 651–661, 2008.

[3] S. Lowe, M. Browne, S. Boudjelas, and M. de Poorter, *100 of the world's worst invasive alien species. A selection from the global invasive species database.*, Invasive Species Specialist Group (ISSG). 2000.

[4] H. Marchante, E. Marchante, and H. Freitas, "Invasion of the Portuguese dune ecosystems by the exotic species Acacia longifolia (Andrews) Willd.: Effects at the community level," in *Plant Invasion: Ecological Threats and Management Solutions*, Backhuys Publishers, pp. 75–85, 2003.

[5] H. Marchante, "Invasion of portuguese dunes by Acacia longifolia : present status and perspective for the future," Tese de Doutoramento, Faculdade de Ciências e Tecnologia da Universidade de Coimbra, 2011.

[6] H. Marchante, "Invasão dos ecossistemas dunares portugueses por Acacia : uma ameaça para a biodiversidade nativa," Tese de Mestrado, Faculdade de Ciências e Tecnologia da Universidade de Coimbra, 2001.

[7] D. C. G. R. Lourenço, "Avaliação de áreas invadidas por espécies de acacia na paisagem protegida da arriba fóssil da Costa de Caparica," Tese de Mestrado, Faculdade de Ciências e Tecnologias da Universidade Nova de Lisboa, 2009.

[8] P. Fernandes, C. Antunes, O. Correia, and C. Máguas, "Do climatic and habitat conditions affect the reproductive success of an invasive tree species? An assessment of the phenology of Acacia longifolia in Portugal," *Plant Ecol.*, vol. 216, pp. 343–355, 2015.

[9] K. G. Rascher, C. Hellmann, C. Máguas, and C. Werner, "Community scale 15N isoscapes: tracing the spatial impact of an exotic N2-fixing invader," *Ecol. Lett.*, vol. 15, pp. 484–491, 2012.

[10] S. Lavergne, N. J. Muenke, and J. Molofsky, "Genome size reduction can trigger rapid phenotypic evolution in invasive plants," *Ann. Bot.*, vol. 105, pp. 109–116, 2010.

[11] M. K. Pandit, S. M. White, and M. J. O. Pocock, "The contrasting effects of genome size, chromosome number and ploidy level on plant invasiveness: a global analysis," *New Phytol.*, vol. 203, pp. 697–703, 2014.

[12] R. V. Gallagher, M. R. Leishman, J. T. Miller, C. Hui, D. M. Richardson, J. Suda, and P. Trávníček, "Invasiveness in introduced Australian acacias: the role of species traits and genome size," *Divers. Distrib.*, vol. 17, pp. 884–897, 2011.

[13] H. Trindade, "Molecular biology of aromatic plants and spices. A review.," *Flavour Fragr. J.*, vol. 25, pp. 272–281, 2010.

[14] L. S. Kumar, "DNA markers in plant improvement: An overview," *Biotechnol. Adv.*, vol. 17, pp. 143–182, 1999.

[15] E. Zietkiewicz, A. Rafalski, and D. Labuda, "Genome Fingerprinting by Simple Sequence Repeat (SSR)-Anchored Polymerase Chain Reaction Amplification," *Genomics*, vol. 20, pp. 176–183, 1994.

[16] I. D. Godwin, E. A. Aitken, and L. W. Smith, "Application of inter simple sequence repeat (ISSR) markers to plant genetics," *Electrophoresis*, vol. 18, pp. 1524–1528, 1997.

[17] M. P. Reddy, N. Sarla, and E. A. Siddiq, "Inter simple sequence repeat (ISSR) polymorphism and its application in plant breeding," *Euphytica*, vol. 128, pp. 9–17, 2002.

[18] B. Bornet and M. Branchard, "Nonanchored Inter Simple Sequence Repeat (ISSR) markers: Reproducible and specific tools for genome fingerprinting," *Plant Mol. Biol. Report.*, vol. 19, pp. 209–215, 2012.

[19] M. Gupta, Y. S. Chyi, J. Romero-Severson, and J. L. Owen, "Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple-sequence repeats," *Theor. Appl. Genet.*, vol. 89, pp. 998–1006, 1994.

[20] M. D. Mendes, A. S. Lima, H. Trindade, A. I. D. Correia, J. G. Barroso, L. G. Pedro, and A. C. Figueiredo, "ISSR molecular characterization and leaf volatiles analysis of Pittosporum undulatum Vent. naturalized in the Azores archipelago (Portugal)," *Ind. Crops Prod.*, vol. 33, pp. 710–719, 2011.

[21] C. C. Josiah, D. O. George, O. M. Eleazar, and W. F. Nyamu, "Genetic diversity in Kenyan populations of Acacia senegal (L.) willd revealed by combined RAPD and ISSR markers," *Afr. J. Biotechnol.*, vol. 7, 2008.

[22] K. A. Selkoe and R. J. Toonen, "Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers," *Ecol. Lett.*, vol. 9, pp. 615–629, 2006.

[23] R. K. Kalia, M. K. Rai, S. Kalia, R. Singh, and A. K. Dhawan, "Microsatellite markers: an overview of the recent progress in plants," *Euphytica*, vol. 177, pp. 309–334, 2010.

[24] T. Guillemaud, L. Broadhurst, I. Legoff, M. Henery, A. Blin, C. Ducatillion, N. Ferrando, and T. Malausa, "Development of 23 polymorphic microsatellite loci in invasive silver wattle, Acacia dealbata (Fabaceae)," *Appl. Plant Sci.*, vol. 3, no. 5, p. 1500018, May 2015.

[25] R. Fredua-Agyeman, D. Adamski, R. J. Liao, C. Morden, and D. Borthakur, "Development and characterization of microsatellite markers for analysis of population differentiation in the tree legume Acacia koa (Fabaceae: Mimosoideae) in the Hawaiian Islands," *Genome*, vol. 51, pp. 1001–1015, 2008.

[26] A. A. Assoumane, A. Vaillant, A. Z. Mayaki, and D. Verhaegen, "Isolation and characterization of microsatellite markers for Acacia senegal (L.) Willd., a multipurpose arid and semi-arid tree," *Mol. Ecol. Resour.*, vol. 9, pp. 1380–1383, 2009.

[27] D. G. Roberts, C. N. Forrest, A. J. Denham, and D. J. Ayre, "Microsatellite markers for vulnerable Australian aridzone Acacias," *Conserv. Genet. Resour.*, vol. 5, pp. 199–201, 2013.

[28] C. N. Forrest, D. G. Roberts, A. J. Denham, and D. J. Ayre, "Microsatellite primers for vulnerable and thriving Acacia (Fabaceae) species from Australia's arid zone," *Appl. Plant Sci.*, vol. 3, p. 1400121, 2015.

[29] D. L. Hartl and A. G. Clark, *Principles of Population Genetics*. Sinauer Associates, 1997.

[30] S. Wright, "Isolation by Distance," *Genetics*, vol. 28, pp. 114–138, 1943.

[31] M. Nei, "Analysis of Gene Diversity in Subdivided Populations," *Proc. Natl. Acad. Sci.*, vol. 70, pp. 3321–3323, 1973.

[32] M. Lynch and B. G. Milligan, "Analysis of population genetic structure with RAPD markers," *Mol. Ecol.*, vol. 3, pp. 91–99, 1994.

[33] S. Wright, "The relation of livestock breeding to theories of evolution," *J. Anim. Sci.*, vol. 46, pp. 1192–1200, 1978.

[34] P. W. Hedrick, "Perspective: Highly Variable Loci and Their Interpretation in Evolution and Conservation," *Evolution*, vol. 53, pp. 313–318, 1999.

[35] J. J. Doyle and J. L. Doyle, "A rapid DNA isolation procedure for small quantities of fresh leaf tissue," *Phytochem. Bull.*, vol. 19, pp. 11–15, 1987.

[36] F. J. Rohlf, *NTSYS-pc: numerical taxonomy and multivariate analysis system*. Applied Biostatistics, 1992.

[37] X. Vekemans, T. Beauwens, M. Lemaire, and I. Roldán-Ruiz, "Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size," *Mol. Ecol.*, vol. 11, pp. 139–151, 2002.

[38] R. Peakall and P. E. Smouse, "genalex 6: genetic analysis in Excel. Population genetic software for teaching and research," *Mol. Ecol. Notes*, vol. 6, pp. 288–295, 2006.

[39] R. Peakall and P. E. Smouse, "GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update," *Bioinformatics*, vol. 28, pp. 2537–2539, 2012.

[40] S. Ueno, J. F. Rodrigues, A. Alves-Pereira, E. R. Pansarin, and E. A. Veasey, "Genetic variability within and among populations of an invasive, exotic orchid," *AoB Plants*, vol. 7, p. plv007, 2015.

[41] S. Y. Zhao, S. G. Sun, C. Dai, R. W. Gituru, J. M. Chen, and Q. F. Wang, "Genetic variation and structure in native and invasive Solidago canadensis populations," *Weed Res.*, vol. 55, pp. 163–172, 2015.

[42] C. L. Taylor and N. P. Barker, "Species limits in Vachellia (Acacia) karroo (Mimosoideae: Leguminoseae): Evidence from automated ISSR DNA 'fingerprinting,'" *South Afr. J. Bot.*, vol. 83, pp. 36–43, 2012.

[43] D. O. Okun, E. U. Kenya, P. O. Oballa, D. W. Odee, and G. M. Muluvi, “Analysis of genetic diversity in Eucalyptus grandis (Hill ex Maiden) seed sources using inter simple sequence repeats (ISSR) molecular markers,” *Afr. J. Biotechnol.*, vol. 7, pp. 2119–2123, 2008.

[44] N. K. Tubić, M. Djan, N. Veličković, G. Anačkov, and D. Obreht, “Microsatellite DNA variation within and among invasive populations of Ambrosia artemisiifolia from the southern Pannonian Plain,” *Weed Res.*, vol. 55, pp. 268–277, 2015.

[45] J. Mulumba, S. Nyakaana, and R. Jamnadass, “Genetic Diversity and Structure of Acacia senegal (L.) Willd. in Uganda,” *Int. J. Bot.*, vol. 8, pp. 107–116, 2012.

[46] S. F. Omondi, E. Kireger, O. G. Dangasuk, B. Chikamai, D. W. Odee, S. Cavers, and D. P. Khasa, “Genetic Diversity and Population Structure of Acacia senegal (L) Willd. in Kenya,” *Trop. Plant Biol.*, vol. 3, pp. 59–70, 2009.

[47] G. D. Thompson, D. U. Bellstedt, D. M. Richardson, J. R. U. Wilson, and J. J. Le Roux, “A tree well travelled: global genetic structure of the invasive tree Acacia saligna,” *J. Biogeogr.*, vol. 42, pp. 305–314, 2015.

[48] K. Weising, H. Nybom, M. Pfenninger, K. Wolff, and W. Meyer, *DNA Fingerprinting in Plants and Fungi*. CRC Press, 1994.

6. Appendices

6.1. Appendix I – Sample Identification

Table 6.1 (left) and Table 6.2 (right): Identification of the samples from Osso da Baleia (left) and Pinheiro da Cruz (right) with a sample code, identification (ID) number and the geographic coordinates of the location.

Sample Code	ID Nº	Geographic Coordinates
OB1	41	40°0'03.29"N 8°54'03.10"W
OB2	42	
OB3	43	
OB4	44	
OB5	45	
OB6	46	
OB7	47	
OB8	48	
OB9	49	
OB10	50	
OB11	51	
OB12	52	
OB13	53	
OB14	54	
OB15	55	
OB16	56	
OB17	57	
OB18	58	
OB19	59	
OB20	60	
OB21	61	
OB22	62	
OB23	63	
OB24	64	
OB25	65	

Sample Code	ID Nº	Geographic Coordinates
PC1	71	38°15'01.38"N 8°45'07.12"W
PC2	72	
PC3	73	
PC4	74	
PC5	75	
PC6	76	
PC7	77	
PC8	78	
PC9	79	
PC10	80	
PC11	81	
PC12	82	
PC13	83	
PC14	84	
PC15	85	
PC16	86	
PC17	87	
PC18	88	
PC19	89	
PC20	90	
PC21	91	
PC22	92	
PC23	93	
PC24	94	
PC25	95	

Table 6.3: Identification of samples from Vila Nova de Milfontes with a sample code, identification (ID) number and the geographic coordinates of the location of each individual (table continues on the next page).

Site	Sample Code	ID Nº	Geographic Coordinates
1960s/70s population	EC	0	External Control: *Acacia saligna*
	MFAnt1	1	37°41'15.55"N, 8°47'31.21"W
	MFAnt2	2	37°41'14.25"N, 8°47'31.91"W
	MFAnt3	3	37°41'12.83"N, 8°47'32.04"W
	MFAnt4	4	37°41'11.37"N, 8°47'31.83"W
	MFAnt5	5	37°41'09.71"N, 8°47'31.24"W
	MFAnt6	6	37°41'08.44"N, 8°47'30.38"W
	MFAnt7	7	37°41'07.05"N, 8°47'29.43"W
	MFAnt9	9	37°41'22.00"N, 8°47'25.92"W
	MFAnt10	10	37°41'23.32"N, 8°47'24.85"W
	MFAnt11	11	37°41'24.41"N, 8°47'23.95"W
	MFAnt12	12	37°41'25.47"N, 8°47'23.12"W
	MFAnt13	13	37°41'26.56"N, 8°47'22.20"W

(Continuation of table 6.3)

Site	Sample Code	ID Nº	Geographic Coordinates
1960s/70s population	MFAnt14	14	37°41'28.13"N, 8°47'21.00"W
	MFAnt15	15	37°41'28.93"N, 8°47'20.12"W
Recent population	MFRec1	16	37°40'53.09"N, 8°46'56.50"W
	MFRec2	17	37°40'56.39"N, 8°46'55.37"W
	MFRec3	18	37°40'50.70"N, 8°46'47.68"W
	MFRec4	19	37°40'49.17"N, 8°46'51.83"W
	MFRec5	20	37°40'50.29"N, 8°46'38.97"W
	MFRec6	21	37°40'58.22"N, 8°46'24.02"W
	MFRec7	22	37°41'01.51"N, 8°46'26.62"W
	MFRec8	23	37°41'04.55"N, 8°46'28.24"W
	MFRec9	24	37°41'25.42"N, 8°46'24.44"W
	MFRec10	25	37°41'28.35"N, 8°46'23.07"W
	MFRec11	26	37°41'33.97"N, 8°46'20.10"W
	MFRec12	27	37°41'32.22"N, 8°46'15.53"W
	MFRec13	28	37°41'29.55"N, 8°46'03.99"W
	MFRec14	29	37°41'02.08"N, 8°46'00.40"W
	MFRec15	30	37°41'01.34"N, 8°45'51.66"W
	MFRec16	31	37°40'57.60"N, 8°45'50.19"W
	MFRec17	32	37°40'50.22"N, 8°45'48.90"W
	MFRec18	33	37°40'46.58"N, 8°45'50.75"W
	MFRec19	34	37°40'41.60"N, 8°45'53.25"W
	MFRec20	35	37°40'32.19"N, 8°45'59.03"W
	MFRec21	36	37°40'43.26"N, 8°46'15.80"W
	MFRec22	37	37°41'02.28"N, 8°47'40.71"W
	MFRec23	38	37°41'03.97"N, 8°47'40.28"W
	MFRec24	30	37°41'05.60"N, 8°47'37.01"W
	MFRec25	40	37°41'00.55"N, 8°47'41.28"W
Coastal Line	MFCL1	101	37°30'27.33"N, 8°27'04.48"W
	MFCL2	102	37°30'26.86"N, 8°27'02.54"W
	MFCL3	103	37°30'26.54"N, 8°27'04.00"W
	MFCL4	114	37°30'42.56"N, 8°26'24.96"W
	MFCL5	104	37°42'14.04"N, 8°46'19.78"W
	MFCL6	105	37°42'20.02"N, 8°46'22.65"W
	MFCL7	106	37°42'29.73"N, 8°46'29.84"W
	MFCL8	107	37°36'01.94"N, 8°48'55.54"W
	MFCL9	108	37°32'34.76"N, 8°47'16.25"W
	MFCL10	109	37°28'37.88"N, 8°47'45.11"W
	MFCL11	110	37°28'35.32"N, 8°44'17.50"W
	MFCL12	111	37°30'47.78"N, 8°43'35.93"W
	MFCL13	112	37°34'28.33"N, 8°44'15.55"W
	MFCL14	113	37°39'39.89"N, 8°45'28.10"W

6.2. Appendix II – DNA Extraction Protocols: Mini-CTAB Method with Ethanol or Isopropanol

The DNA was extracted with ethanol using a method developed by Doyle & Doyle (1987)[35] modified by Weising *et al.* (1995)[48] and adapted to *Acacia longifolia* as follows (the differences between the ethanol and isopropanol extraction protocols are denoted with "•"):

1. Pre-heat CTAB extraction buffer (2% (w/v) CTAB, 1.4 M NaCl, 20 mM EDTA, 100 mM Tris-HCl pH 8.0, PVP-40 1-2%) at 65 °C;
2. Maceration of approximately half a phyllode (50-100 mg) of *A. longifolia* with liquid nitrogen and transference of the biological material to a 2 mL microtube;
3. Addition of 1 mL of CTAB extraction buffer, 100 µL of chloroform:isoamyl alcohol 24:1 and 4 µL of β-mercaptoethanol;
    - For extraction with isopropanol, addition of 600 µL of CTAB extraction buffer, 600 µL of chloroform:isoamyl alcohol 24:1 and 6 µL of β-mercaptoethanol.
4. Incubation at 65 °C for 30 min;
5. Filling of the remaining tube volume with chloroform:isoamyl alcohol 24:1 and centrifugation at 15000xg for 10 min at room temperature;
    - For extraction with isopropanol, addition of 600 µL of chloroform:isoamyl alcohol 24:1 and centrifuge with same conditions.
6. Transference of the light-green, translucent supernatant at the top of the tube to another clean 2 mL microtube and addition of cold (-20°C) 100% ethanol until the tube is filled. The DNA precipitates in thin, white fibers visible to the naked eye;
    - DNA precipitation with 0.6 volumes of isopropanol in the same conditions.
7. Incubation at -20 °C for 2h;
8. Centrifugation at 10500xg for 20 min at 4°C;
9. Removal of the supernatant and addition of 500 µL of the "Washing Solution" (ammonium acetate 0,01 M in 76% ethanol);
10. Centrifugation at 5000xg for 10 min at 4°C;
11. Removal of the supernatant and air drying of the white pellets for 30-40 min (dependent on the size of the pellet);
12. Resuspension in 200-400 µL 1x TE buffer (dependent on the size of the pellet).

6.3. Appendix III – Inter-Simple Sequence Repeats (ISSRs) Primers

Table 6.4: List ISSR primers and corresponding annealing temperatures. TD - touch-down PCR cycle**.** [1]Primers selected for analysis of samples from Vila Nova de Milfontes (1960s/70s and recent sites), Osso da Baleia and Pinheiro da Cruz. [2]Primers selected for analysis of samples from Vila Nova de Milfontes (all sites).

Primer	Sequence	Annealing temperature (°C)
807	$(AT)_8T$	52
808[1,2]	AGAGAGAGAGAGAGAGC	52
810[1,2]	GTGTGTGTGTGTGTGTCA	52
812	$(GA)_8A$	53
813[1,2]	CTCTCTCTCTCTCTCTT	52
817[1]	CACACACACACACACAA	52
823[2]	$(TC)_8C$	54
825[1]	$(AC)_8T$	54
826[1]	$(AC)_8C$	58
827[1,2]	$(AC)_8G$	58
834	$(AG)_8YT$	59
835	$(AG)_8YC$	59
836[1,2]	$(AG)_8YA$	54
840[1,2]	$(GA)_8YT$	TD 57/54
841	$(GA)_8YC$	54
846[1,2]	$(CA)_8RT$	TD 57/53
849[1,2]	GAGAGAGAGAGAGAGAT	52
857	$(AC)_8YG$	52
858	$(TG)_8RT$	59
862	$(AGC)_6$	55
864	ATGATGATGATGATGATG	52
866	$(CTC)_6$	55
868[1,2]	$(GAA)_6$	51
880[1]	$GGA(GAG)_2AGGAGA$	54
892	TAG ATC TGA TAT CTG AAT TCC C	54
895	AGA GTT GGT AGC TCT TGA TC	55
899	CAT GGT GTT GGT CAT TGT TCC A	58
900	ACT TCC CCA CAG GTT AAC ACA	57
901[1,2]	DHB(CGA)5x	TD 57/52
903[1]	BDB(CAC)5x	TD 57/52
904[1]	$DDC(CAC)_5$	TD 57/52

6.4. Appendix IV – Simple Sequence Repeats (SSRs) or Microsatellite Pairs of Primers

Table 6.5: List of tested SSR primers, corresponding sequence and reference.

Primer	Sequence (5'-3')	References
BBY8P	Fw: TTGGCAAATCCGCACAGTC	[28]
	Rv: TGCCATCGCAACATATAGCTTC	
BA1R8	Fw: GGTGCTTTTCCCCACCTTC	[28]
	Rv: TCTCGCTTTTCATGTGCAAG	
CIDYF	Fw: CACACTTATGGGATGGGTTGC	[27], [28]
	Rv: AGCTAAGGAAAGTGTACGGGAAT	
CPUH4	Fw: AGATGCATTGACTGAGACGG	[28]
	Rv: CGAATGAAGGAGATTTATGAAGAGAC	
AV9GR	Fw: CCAACGACAGTGGGCAGTC	[28]
	Rv: CTCCGGTGTTAGCAAAGGC	
BVWHY	Fw: TCCTACTTCCCCAACACGC	[28]
	Rv: ACAAGCAGCCATTGGAAGG	
APZIZ	Fw: ACACTACACTCACAACACACAC	[28]
	Rv: ACACGGTTTGCTTGGCTTG	
AO12C	Fw: AAAACAAGAGAAGAGGACATGC	[28]
	Rv: TCGTAGAAACGACACGAAACG	
CU0EQ	Fw: ACCACCATCTTCACCTCCAC	[28]
	Rv: TCCGGCGTTTCCAACTAAC	
ACPU7	Fw: GTTCTACGGCTAGATGGTGC	[28]
	Rv: TGTCATACGGCCTCACAAAG	
DCLOC	Fw: CAACTTGTGATTAAAGTCCACGG	[27]
	Rv: TGTGTTGAGACTTTGTGCTG	

Table 6.6: List of selected SSR primers and corresponding fluorescent label, expected fragment size and GenBank number.

Primer	Sequence (5'-3')	Fluorescent Label 5'	Fragment size (bp)	GenBank number
CPUH4	Fw: AGATGCATTGACTGAGACGG	ATTO-550	112-115	KF776129
	Rv: CGAATGAAGGAGATTTATGAAGAGAC			
APZIZ	Fw: ACACTACACTCACAACACACAC	6-FAM	222-250	KF776135
	Rv: ACACGGTTTGCTTGGCTTG			
DCLOC	Fw: CAACTTGTGATTAAAGTCCACGG	HEX	128-160	-----
	Rv: TGTGTTGAGACTTTGTGCTG			